Efficient asymptotic preserving deterministic methods for the Boltzmann equation

Lorenzo Pareschi[*] Giovanni Russo[†]

April 1, 2011

Contents

[*] Department of Mathematics, University of Ferrara, Ferrara, Italy
[†] Department of Mathematics and Computer Science, University of Catania, Catania, Italy

Report Documentation Page	*Form Approved* *OMB No. 0704-0188*

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **JAN 2011**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Efficient asymptotic preserving deterministic methods for the Boltzmann equation**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Mathematics, University of Ferrara, Ferrara, Italy**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S)
	11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES

See also ADA579248. Models and Computational Methods for Rarefied Flows (Modeles et methodes de calcul des coulements de gaz rarefies). RTO-EN-AVT-194

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**SAR**	**76**	

Introduction

In this lecture notes we review some recent results concerning the numerical solution of nonlinear collisional kinetic equation. The most well-known example is represented by the Boltzmann equation of rarefied gas dynamics (Cercignani, 1988; Cercignani et al., 1994). Besides other classical examples, like the Landau equation of plasma physics (Landau, 1981), kinetic equations play an important role in modelling granular gases (Bobylev et al., 2000), charged particles in semiconductors (Markowich et al., 1989), neutron transport (Jin et al., 2000) and quantum gases (Escobedo et al., 2003b). More recently applications of kinetic equations have been considered for car traffic flows (Klar and Wegener, 1997), chemotactical movements (Chalub et al., 2004), tumor immune cells competition (Bellomo and Bellouquid, 2004), coagulation-fragmentation processes (Escobedo et al., 2003a), population dynamics (Desvillettes et al., 2004), market economies (Cordier et al., 2005), supply chains (Armbruster et al., 2007), flocking dynamics (Ha and Tadmor, 2008) and many other. For a recent introduction to the Boltzmann equation and related kinetic equations we refer the reader to Degond et al. (2004); Villani (2002), recent applications to biology and socio-economy can be found in Naldi et al. (2010).

Although the scope of our insights is wider, here we will focus mainly on the classical Boltzmann equation of rarefied gas dynamics. This is motivated not only by its relevance for applications but also because it contains all major difficulties present in other kinetic models and represents the most challenging case for the development of numerical schemes.

Approximate methods of solution for the Boltzmann equation have a long history tracing back to Hilbert, Chapmann and Enskog (Cercignani, 1988) at the beginning of the last century. The mathematical difficulties related to the Boltzmann equation make it extremely difficult, if not impossible, the determination of analytic solutions in most physically relevant situations. Only in recent years, starting in the 70s with the pioneering works by Chorin (1972) and Sod (1977), the problem has been tackled numerically with particular care to accuracy and computational cost.

Most of the difficulties are due to the multidimensional structure of the collisional integral, since the integration runs on a highly-dimensional unflat manifold. In addition the numerical integration requires great care since the collision integral is at the basis of the macroscopic properties of the equation. Further difficulties are represented by the presence of stiffness, like the case of small mean free path (Gabetta et al., 1997) or the case of large velocities (Filbet and Pareschi, 2003).

For such reasons realistic numerical simulations are based on Monte-Carlo techniques. The most famous examples are the Direct Simulation Monte-Carlo (DSMC) methods by Bird (Bird, 1994) and by Nanbu (Nanbu, 1980). These methods guarantee efficiency and preservation of the main physical properties. However, avoiding statistical fluctuations in the results becomes extremely expensive in presence of non-stationary flows or close to continuum regimes.

Among deterministic approximations one of the most popular method is represented by the so called Discrete Velocity Models (DVM) of the Boltzmann equation. These methods (Martin et al., 1992; Rogier and Schneider, 1994; Bobylev et al., 1995; Buet, 1996; Panferov and Heintz, 2002) are based on a Cartesian grid in velocity and on a discrete collision mechanism on the points of the grid that preserves the main physical properties. Unfortunately DVM are not competitive with Monte Carlo methods in terms of computational cost (typically $O(n^{(2d_v+1)/d_v})$, where n is the total number of discretization parameters in velocity and d_v is the dimension of the velocity space) and their accuracy seems to be at most first order in velocity (Palczewski et al., 1997; Palczewski and Schneider, 1998; Panferov and Heintz, 2002).

Another important class of numerical methods is based on the use of spectral techniques in the velocity space. The methods were first derived in Pareschi and Perthame (1996), inspired by previous works on the use of Fourier transform techniques (see Bobylev (1988) for instance). The numerical method is based on approximating the distribution function by a periodic function in velocity space, and on its representation by Fourier series. The resulting scheme can be evaluated with a computational cost of $O(n^2)$.

The method was further developed in Pareschi and Russo (2000b,c) where evolution equations for the Fourier modes were explicitly derived and spectral accuracy of the method was proved. Strictly speaking these methods are not conservative, since they preserve mass, whereas momentum and energy are approximated with spectral accuracy. This trade off between accuracy and conservations seems to be an unavoidable compromise in the development of numerical schemes for the Boltzmann equation.

Recently in Mouhot and Pareschi (2006, 2004); Filbet et al. (2006), using a suitable representation of the collision operator, the computational cost of spectral methods has been reduced from $O(n^2)$ to $O(n \log_2 n)$ without loosing the spectral accuracy thus making the methods competitive with Monte Carlo. These fast algorithms are restricted to a certain class of particle interactions including pseudo-Maxwell molecules (for $d_v = 2$) and hard spheres (for $d_v = 3$). This kind of approach has been extended recently to construct fast algorithms for DVM models (Mouhot and Pareschi, 2011). Another class of fast solvers for the case of radially symmetric distribution functions has been constructed in Markowich and Pareschi (2005).

We recall here that the spectral method has been applied also to non homogeneous situations (Filbet and Russo, 2003), to the Landau equation (Filbet and Pareschi, 2003; Pareschi et al., 2000, 2003), where fast algorithms can be readily derived, to the case of granular gases (Naldi et al., 2003; Filbet et al., 2005) and more recently to the case of a quantum gas (Filbet et al., 2011). Let us mention that algorithms based on a Fourier transform approximation of the distribution function have been constructed in Bobylev and Rjasanow (1997, 1999) and more recently in Gamba and Tharkabhushanam (2009, 2010). Other fast algorithms for kinetic equations can be found in Buet et al. (1997); Lemou (1998).

An additional problem in the numerical solution of the Boltzmann equation is the time step restriction in regions close to the fluid dynamic limit, i.e. for very small Knudsen number. In such a cases, in fact, the mean collision time is so small that an explicit solver would require a very small time step, thus degrading the performance of the method.

If the Knudsen number is uniformly small, then one usually does not need a kinetic treatment, and the gas can be very well described by the Euler or Navier-Stokes equations. There are cases, however, in which the local Knudsen number varies over several orders of magnitude. Several authors have tackled the problem in the past, and there is a large literature on the subject (see Bennoune et al. (2008); Caflisch et al. (1997); Degond et al. (2005); Gabetta et al. (1997); Filbet and Jin (2010); Tiwari and Klar (1998) and the references therein).

One possibility is to resort to the so called *domain decomposition* techniques: one could divide the computational domain into two complementary domains, let us denote them by A[erodynamics] and B[oltzmann]; in A the gas is well described by either Euler or Navier-Stokes equations, while in B the gas needs a kinetic description. In some cases the subdivision into such two subdomains may be known *a priori*, for example from the geometry of the domain of from previous approximate calculations of the flow, but in most cases the two regions are themselves unknown, and therefore they have to be computed and evolved as part of the solution. Examples of Euler-Boltzmann coupling can be found, for example, in Bourgat et al. (1992); Tiwari and Klar (1998); Tiwari (1998), while Navier-Stokes-Boltzmann coupling is considered in Bourgat et al. (1996). More recent works, including hybrid methods, can be found in Schwartzentruber et al. (2007); Degond et al. (2007); Dimarco and Pareschi (2010b).

The coupling between aerodynamics and kinetic description is very appealing, because one could gain maximum efficiency by treating with continuum equations the region in which a kinetic description is not strictly necessary. However, it presents several difficulties, and it is still an open problem to asses what is the best coupling strategy.

In these notes we shall consider system which is treated by kinetic equations in the whole computational domain. We shall discuss what strategies can be used to treat regions with small Knudsen number avoiding unnecessary restrictions on the time step. When the mean collision time is much smaller that typical length scales of the problem, the Boltzmann equation becomes *stiff*. Usually stiff systems governed by time dependent equations can be effectively treated by implicit schemes. The interested reader may consult the monograph by Hairer and Wanner on numerical solution of stiff problems (Hairer and Wanner, 1996). A direct time discretization of the Boltzmann equation seems not possible in such stiff regimes due to the high dimensionality and the nonlinearity of the collision operator which makes unpractical the use of implicit solvers.

Here we present two classes of methods which avoid the solution of systems of nonlinear equations. The first one is based on operator splitting approach combined with the use of exponential techniques, and will be applied to the Boltzmann equation itself (see Gabetta et al. (1997); Dimarco and Pareschi (2010a)). The second class is based on non splitting approaches. We focus our attention on implicit semilagrangian schemes (see Santagati (2007); Russo and Santagati (2011)) and Implicit-Explicit Runge-Kutta methods (see Pareschi and Russo (2005); Pieraccini and Puppo (2007)); such methods will be constructed for the BGK model of the Boltzmann equation and then their extension to the full Boltzmann equation discussed (Filbet and Jin, 2010; Dimarco and Pareschi, 2011). We refer the reader to Bennoune et al. (2008) for a related approach.

In addition to other deterministic methods that will not be discussed in the notes, such as finite difference methods (Ohwada, 1993; Sone et al., 1989), there are several important aspects that we shall not be able to discuss in the present paper, namely:

- Numerical treatment of boundary conditions. In most applications reflective or absorbing boundary conditions should be considered, or a suitable combination of the two. However a detailed treatment of boundary conditions depends on the geometry of the domain, and on the detail of the space discretization (Carrillo et al., 2006).

- Large departure from (global) equilibrium. One of the weak points of deterministic methods based on a fixed grid in velocity space is that a huge grid in velocity would be necessary to treat flows with a large variation of macroscopic mean velocity and temperature. Effective schemes for the treatment of such cases have to be based on the use of grid in velocity domain which change from one space location to another (Filbet and Russo, 2006; Heintz et al., 2008).

- Multiple space regimes. Most realistic flows present strong non homogeneities, with small regions of large gradients in the moments. Even the region requiring kinetic treatment may present several space scales. An effective treatment of such problems would require adaptivity of the grid in space (Cai and Li, 2010).
- Effective techniques for stationary flows. These notes provide a review of deterministic methods for the time dependent Boltzmann equation. In fact, thanks to averaging procedures, the stationary case is usually computed efficiently with Monte Carlo methods (Bird, 1994). However, one may be interested in accurate deterministic computations of the stationary solutions, which may be treated by schemes aimed to capture the stationary state (Greenberg and Leroux, 1996; Botchorishvili et al., 2003).
- Diffusion limits. In several circumstances the space-time scaling of the kinetic equation leads asymptotically to the corresponding diffusion system. These problems present additional difficulties compare to the standard fluid scaling since also the transport terms are strongly stiff (Jin et al., 2000; Lemou and Mieussens, 2008).

1 The Boltzmann equation

1.1 The model

The model is characterized by a density function $f(x, v, t)$ describing the time evolution of a monoatomic rarefied gas of particles which move with velocity $v \in I\!R^3$ in the position $x \in \Omega \subset I\!R^3$ at time $t > 0$ which satisfies the Boltzmann equation (Cercignani, 1988; Cercignani et al., 1994)

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \frac{1}{\varepsilon} Q(f, f), \tag{1}$$

with initial data

$$f(x, v, 0) = f_0(x, v). \tag{2}$$

The parameter $\varepsilon > 0$ is called *Knudsen number* and is proportional to the mean free path between collisions. The bilinear collision operator $Q(f, f)$ which describes the binary collisions of the particles acts over the velocity variable only

$$Q(f, f)(v) = \int_{I\!R^3} \int_{\mathbb{S}^2} B(v, v_*, \omega)[f(v')f(v'_*) - f(v)f(v_*)]\, d\omega\, dv_*. \tag{3}$$

In the above expression, ω is a unit vector of the sphere $\mathbb{S}^2$ and (v', v'_*) represent the collisional velocities associated with (v, v_*). The collisional velocities satisfy microscopic momentum and energy conservation

$$v' + v'_* = v + v_*, \quad |v'|^2 + |v'_*|^2 = |v|^2 + |v_*|^2. \tag{4}$$

The above system of algebraic equations has the following parametrized solution

$$v' = \frac{1}{2}(v + v_* + |v - v_*|\omega), \quad v'_* = \frac{1}{2}(v + v_* - |v - v_*|\omega) \tag{5}$$

where $v - v_*$ is the relative velocity.

The collision kernel $B(v, v_*, \omega)$ is a nonnegative function which characterizes the details of the binary interactions and depends only on $|v - v_*|$ and the scattering angle θ between relative velocities $v - v_*$ and $v' - v'_* = |v - v_*|\omega$

$$\cos\theta = \frac{(v - v_*) \cdot \omega}{|v - v_*|}.$$

The kernel has the from

$$B(v, v_*, \omega) = |v - v_*|\sigma(|v - v_*|, \cos\theta), \tag{6}$$

where the function σ is the *scattering cross-section.*

Example 1

- *In the* hard sphere model *the particles are assumed to be ideally elastic spheres of diameter $d > 0$ and thus*

$$\sigma(|v - v_*|, \cos\theta) = \frac{d^2}{4}, \quad B(v, v_*, \omega) = \frac{d^2}{4}|v - v_*|, \tag{7}$$

  since the total cross section is $\pi d^2 = 4\pi(d^2/4)$.

- *In the case of inverse k-th power forces between particles, the kernel has the form*

$$\sigma(|v - v_*|, \cos\theta) = b_\alpha(\cos\theta)|v - v_*|^{\alpha-1}, \quad B(v, v_*, \omega) = b_\alpha(\cos\theta)|v - v_*|^\alpha, \tag{8}$$

  with $\alpha = (k - 5)/(k - 1)$. For $k > 5$ we have hard potentials*, for $k < 5$ we have* soft potentials*.*

- *The special situation $k = 5$ gives the* Maxellian model *with*

$$B(v, v_*, \omega) = b_0(\cos\theta). \tag{9}$$

- *For numerical purposes, a widely used model is the* Variable Hard Sphere *(VHS) model, corresponding to $b_\alpha(\cos\theta) = C_\alpha$, where C_α is a positive constant, and hence*

$$\sigma(|v - v_*|, \cos\theta) = C_\alpha|v - v_*|^{\alpha-1}, \quad B(v, v_*, \omega) = C_\alpha|v - v_*|^\alpha. \tag{10}$$

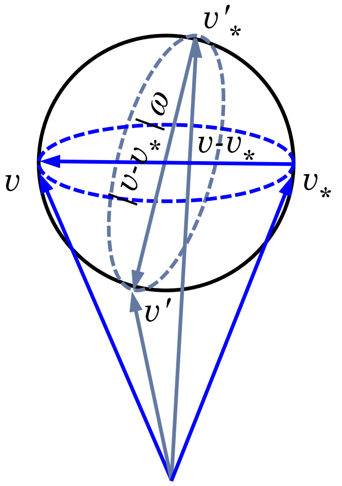


Figure 1: The collision sphere

The collision integral $Q(f,f)$ can be written in different equivalent forms, according to the parametrization used for the collisional velocities. Using the identity

$$\int_{\mathbb{S}^2} (u\cdot n)_+ \phi(n(u\cdot n))\,dn = \frac{|u|}{4}\int_{\mathbb{S}^2}\phi\left(\frac{u-|u|\omega}{2}\right)d\omega \tag{11}$$

obtained by the transformation $\omega = e - 2(e\cdot n)n$, we get the frequently used form

$$Q(f,f)(v) \;=\; \int_{I\!R^3}\int_{\mathbb{S}^2}\tilde{B}(v,v_*,\omega)[f(v')f(v'_*)-f(v)f(v_*)]\,d\omega\,dv_* \tag{12}$$

with

$$v' = v-((v-v_*)\cdot\omega)\omega, \quad v'_* = v_* + ((v-v_*)\cdot\omega)\omega, \tag{13}$$

and

$$\tilde{B}(v,v_*,\omega) = 2|v-v_*||\cos\theta|\sigma(|v-v_*|,1-2|\cos\theta|). \tag{14}$$

The hard sphere case corresponds to

$$\tilde{B}(v,v_*,\omega) = \frac{d^2}{2}|v-v_*||\cos\theta|, \tag{15}$$

whereas the Maxwellian molecules case gives

$$\tilde{B}(v,v_*,\omega) = 2|\cos\theta|b_0(\cos\theta). \tag{16}$$

Remark 1

For the Maxwellian case the collision kernel $B(v,v_*,\omega)$ is independent of the relative velocity. This case has been widely studied theoretically, in particular exact analytic solutions can be found in the space homogeneous case where $f=f(v,t)$ (Bobylev, 1975).

A simplified one-dimensional space homogeneous Maxwell model is given by the Kac equation *(Kac, 1957). It reads*

$$\frac{\partial f}{\partial t} = \int_{\mathbb{R}} \int_0^{2\pi} \frac{1}{2\pi} [f(v'_*) f(v') - f(v) f(v_*)] \, d\theta \, dv_* \tag{17}$$

where the collisional velocities are characterized by rotations in the collisional plane

$$v'_* = v \cos\theta - v_* \sin\theta, \quad v'_* = v \sin\theta + v_* \cos\theta. \tag{18}$$

For this model we have only microscopic conservation of energy $(v')^2 + (v'_*)^2 = v^2 + v_*^2$.

1.2 Physical properties

During the evolution process, the collision operator preserves mass, momentum and energy, i.e.,

$$\int_{\mathbb{R}^3} Q(f,f)\phi(v) \, dv = 0, \quad \phi(v) = 1, v^x, v^y, v^z, |v|^2, \tag{19}$$

and in addition it satisfies Boltzmann's well-known H-theorem

$$\int_{\mathbb{R}^3} Q(f,f) \ln(f(v)) dv \leq 0. \tag{20}$$

The above properties are a consequence of the following identity that can be easily proved for any test function $\phi(v)$

$$\int_{\mathbb{R}^3} Q(f,f)\phi(v) \, dv = -\frac{1}{4} \int_{\mathbb{R}^6} \int_{\mathbb{S}^2} B(v, v_*, \omega)[f' f'_* - f f_*][\phi' + \phi'_* - \phi - \phi_*] \, d\omega \, dv_* \, dv.$$

where we have omitted the explicit dependence from v, v_*, v', v'_* to simplify the expression.

In order to prove this identity we used the micro-reversibility property $B(v, v_*, \omega) = B(v_*, v, \omega)$ and the fact that the Jacobian of the transformation $(v, v_*) \leftrightarrow (v', v'_*)$ is equal to 1.

A function ϕ such that

$$\phi(v') + \phi(v'_*) - \phi(v) - \phi(v_*) = 0$$

is called a *collision invariant*. It can be shown that a continuous function ϕ is a collision invariant if and only if $\phi \in \text{span}\{1, v, |v|^2\}$ or equivalentely

$$\phi(v) = a + b \cdot v + c|v|^2, \quad a, c \in \mathbb{R}, \quad b \in \mathbb{R}^3.$$

Assuming f strictly positive, for $\phi(v) = \ln(f(v))$ we obtain

$$\begin{aligned}
&\int_{\mathbb{R}^3} Q(f,f) \ln(f) dv \\
&= -\frac{1}{4} \int_{\mathbb{R}^6} \int_{\mathbb{S}^2} B(v, v_*, \omega)[f' f'_* - f f_*][\ln(f') + \ln(f'_*) - \ln(f) - \ln(f_*)] \, d\omega \, dv_* \, dv \\
&= -\frac{1}{4} \int_{\mathbb{R}^6} \int_{\mathbb{S}^2} B(v, v_*, \omega)[f' f'_* - f f_*] \ln\left(\frac{f' f'_*}{f f_*}\right) \, d\omega \, dv_* \, dv \leq 0,
\end{aligned}$$

since the function $z(x, y) = (x - y)\ln(x/y) \geq 0$ and $z(x, y) = 0$ only if $x = y$.

In particular, the equality holds only if $\ln(f)$ is a collision invariant, that is

$$f = \exp(a + b \cdot v + c|v|^2), \quad c < 0.$$

If we define the density, mean velocity and temperature of the gas by

$$\rho = \int_{\mathbb{R}^3} f\, dv, \qquad u = \frac{1}{\rho}\int_{\mathbb{R}^3} vf\, dv, \qquad T = \frac{1}{3R\rho}\int_{\mathbb{R}^3}[v - u]^2 f\, dv, \tag{21}$$

we obtain that the distribution function has the form of a locally *Maxwellian distribution*

$$f(v, t) = M(\rho, u, T)(v, t) = \frac{\rho}{(2\pi RT)^{3/2}}\exp\left(-\frac{|u - v|^2}{2RT}\right).$$

The constant $R = K_B/m$ is called the gas constant, K_B is the Boltzmann constant and m the mass of a particle. Boltzmann's H-theorem implies that any equilibrium distribution function, i.e. any function f for which $Q(f, f) = 0$, has the form of a locally Maxwellian distribution.

If we define the *H-function*

$$H(f) = \int_{\mathbb{R}^3} f\ln(f)\, dv,$$

we obtain immediately the inequality

$$\frac{dH(f)}{dt} = \int_{\mathbb{R}^3} Q(f, f)\ln(f)\, dv \leq 0. \tag{22}$$

Thus the H-function is monotonically decreasing until f reaches the equilibrium Maxwellian state for which we have

$$H(M) = \rho\left(\ln\left(\frac{\rho}{(2\pi RT)^{3/2}}\right) - \frac{3}{2}\right).$$

1.3 Fluid limit

If we multiply the Boltzmann equation by its collision invariants and integrate the result in velocity space we obtain

$$\frac{\partial}{\partial t}\int_{\mathbb{R}^3} f\phi(v)\, dv + \nabla_x\left(\int_{\mathbb{R}^3} vf\phi(v)\, dv\right) = 0, \quad \phi(v) = 1, v_1, v_2, v_3, |v|^2.$$

These equations describe the balance of mass, momentum and energy. The system of five equations is not closed since it involves higher order moments of the distribution function f.

As $\varepsilon \to 0$, from (1) we have formally $Q(f, f) \to 0$, and thus f approaches the local Maxwellian. In this case the higher order moments of the distribution function can be computed as function of ρ, u, and T and we obtain the closed system of *compressible Euler equations*

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho u) = 0 \\
&\frac{\partial \rho u}{\partial t} + \nabla_x \cdot (\rho u \otimes u + p) = 0 \\
&\frac{\partial E}{\partial t} + \nabla_x \cdot (Eu + pu) = 0 \\
&p = \rho T, \quad E = \frac{3}{2}\rho T + \frac{1}{2}\rho u^2
\end{aligned}$$

where p is the gas pressure and $\otimes$ denotes the tensor product.

The rigorous passage from the Boltzmann equation to the compressible Euler equations has been investigated by several authors. Among them we mention references Caflisch (1980); Nishida (1978). Higher order fluid models, such as the Navier-Stokes model, can be considered using the Chapmann-Enskog and the Hilbert expansions. We refer to Levermore (1996) for a mathematical setting of the problem and to Golse and Saint-Raymond (2004) for recent theoretical results.

1.4 Boundary conditions

The Boltzmann equation is complemented with the boundary conditions in space for $v \cdot n \geq 0$ and $x \in \partial\Omega$, where n denotes the unit normal, pointing inside the domain Ω. Usually the boundary represents the surface of a solid object (an obstacle or a container). The particles of the gas that hit the surface interact with the atoms of the object and are reflected back into the domain Ω.

Mathematically, such boundary conditions are modelled by an expression of the form (Cercignani, 1988)

$$|v \cdot n| f(x, v, t) = \int_{v_* \cdot n < 0} |v_* \cdot n(x)| K(v_* \to v, x, t) f(x, v_*, t)\, dv_*. \tag{23}$$

This is the so-called *reflective boundary condition* on $\partial\Omega$.

The ingoing flux is defined in terms of the outgoing flux modified by a given boundary kernel K. This boundary kernel is such that positivity and mass conservation at the

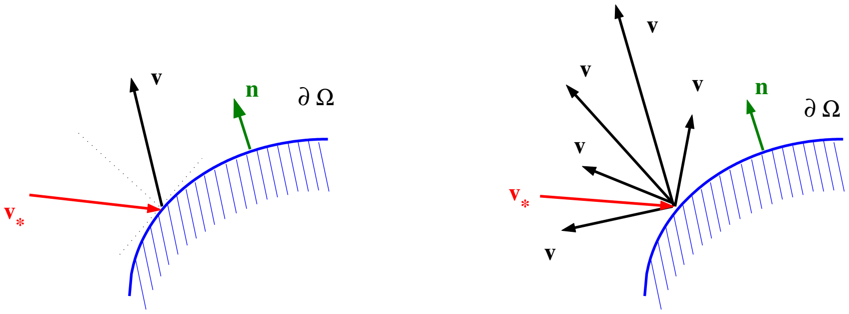


Figure 2: Reflection and diffusion at the solid boundary

boundaries are guaranteed,

$$K(v_* \to v, x, t) \geq 0, \qquad \int_{v\cdot n(x)\geq 0} K(v_* \to v, x, t)\, dv = 1.$$

Commonly used reflecting boundary conditions are the so-called Maxwell's conditions. From a physical point of view, one assumes that a fraction α of molecules is absorbed by the wall and then re-emitted with the velocities corresponding to those in a still gas at the temperature of the wall, while the remaining fraction $(1-\alpha)$ is specularly reflected.

This is equivalent to impose for the ingoing velocities

$$f(x, v, t) = (1-\alpha)Rf(x, v, t) + \alpha Mf(x, v, t), \tag{24}$$

in which $x \in \partial\Omega, \quad v \cdot n(x) \geq 0$. The coefficient α, with $0 \leq \alpha \leq 1$, is called the *accommodation coefficient* and

$$Rf(x, v, t) = f(x, v - 2n(n \cdot v), t), \quad Mf(x, v, t) = \mu(x, t)M_w(v). \tag{25}$$

If we denote by T_w the temperature of the solid boundary, M_w is given by

$$M_w(v) = \exp(-\frac{v^2}{2RT_w}),$$

and the value of μ is determined by mass conservation at the surface of the wall

$$\mu(x, t) \int_{v\cdot n \geq 0} M_w(v)|v \cdot n| dv = \int_{v\cdot n<0} f(x, v, t)|v \cdot n| dv. \tag{26}$$

For $\alpha = 0$ (specular reflection) the re-emitted molecules have the same flow of mass, temperature and tangential momentum of the incoming molecules, while for $\alpha = 1$ (full accommodation) the re-emitted molecules have completely lost memory of the incoming molecules, except for conservation of the number of molecules.

More complex boundary conditions for rarefied gas dynamics (RGD) can be imposed using the boundary conditions of Cercignani and Lampis (Cercignani and Lampis, 1971). These can be written as

$$f(x, v, t) = \int P(v, v')f(x, v', t)dv' \tag{27}$$

where

$$P(v, v') = (2v/\alpha)I'(2(1-\alpha)1/2vv'/\alpha)\exp\left(v^2 - (1-\alpha)v'^2\right)/\alpha \tag{28}$$

in which v and v' are the normal components of the outgoing and incoming velocities respectively, and I' is the modified Bessel function. This satisfies the reciprocity (detailed balance) condition

$$vP(-v', -v)M_w(v) = -v'P(v, v')M_w(v'). \tag{29}$$

A consequence of the reciprocity condition is that the Maxwellian distribution M_w is preserved by this boundary condition.

In the case of *inflow boundary conditions*, one assumes that the distribution function of the particles entering the domain is known, i.e.

$$f(x,v,t)=g(v,t),\quad x\in\partial\Omega,\quad v\cdot n>0,$$

A typical example of such condition is used in shock wave calculations, where one assumes that the distribution function at the boundary of the computational domain is a Maxwellian $M(v)$ and that the incoming flux of particles at the boundary is distributed according to the Maxwellian flux $(v\cdot n)M(v)$, $v\cdot n>0$.

1.5 Other collision operators

1.5.1 BGK models

A simplified model Boltzmann equation is the so-called BGK model introduced by Bhatnagar et al. (1954). In this model the collision operator is replaced by a relaxation operator of the form

$$Q_{\mathrm{BGK}}(f,f)(v)=\frac{1}{\tau}(M[f]-f), \tag{30}$$

where $M[f]=M(v;\{\rho,u,T\})$ is the local Maxwellian computed by the moments of the distribution function f

$$M(v;\{\rho,u,T\})=\frac{\rho}{(2\pi RT)^{3/2}}\exp\left(-\frac{|v-u|^2}{2RT}\right). \tag{31}$$

where $\rho=\rho(x,t)$, $u=u(x,t)$ and $T=T(x,t)$ denote the macroscopic fields, namely: density, mean velocity and temperature, corresponding to the function f.

Conservation of mass, momentum and energy as well as Boltzmann H-theorem are readily satisfied. The equilibrium solutions are clearly Maxwellians

$$Q_{\mathrm{BGK}}(f,f)=0\Leftrightarrow f=M[f].$$

The relaxation time τ is in general inversely proportional to the density, and depends on the temperature:

$$\tau^{-1}=A(T)\rho$$

Numerical computations, as well as the analytic theory, for such model are much simpler then for the full Boltzmann equation.

Furthermore, as a consequence of conservation of mass, momentum and energy, in the fluid dynamic limit the moments (ρ,
$rhou$, and E), i.e. mass density, momentum density, and total energy density, satisfy the compressible Euler equations for a monoatomic gas, therefore the model describes the correct fluid dynamic limit.

But in the Chapman-Enskog expansion, the transport coefficients obtained at the Navier-Stokes level are not satisfactory. The relaxation time could be adjusted so that at the Navier-Stokes level the model provides the correct value of one transport coefficient, say the viscosity. However, the Prandtl number P_r (the ratio between heat conductivity and viscosity) is equal to 1. For most gases, we have $P_r<1$. In particular, the hard-sphere model for a monoatomic gas leads to a Prandtl number very close to 2/3, therefore only one transport coefficient can be correct, but not both. The correct Prandtl number can be recovered using more sophisticated BGK models, as the velocity dependent collision frequency BGK models and the Ellipsoidal Statistical BGK (ES-BGK) models (Bouchut and Perthame, 1993; Holway, 1966).

1.5.2 Landau models

The Landau model (Landau, 1981) is a common kinetic model in plasma physics characterized by the following collision operator

$$Q_L(f,f)(v) = \nabla_v \cdot \int_{\mathbb{R}^d} A(v-v_*)[f(v_*)\nabla_v f(v) - f(v)\nabla_{v_*} f(v_*)]\, dv_*$$

where $A(z) = \Psi(|z|)\Pi(z)$ is a $d \times d$ nonnegative symmetric matrix and $\Pi(z) = (\pi_{ij}(z))$ is the orthogonal projection upon the space orthogonal to z,

$$\pi_{ij}(z) = \left(\delta_{ij} - \frac{z_i z_j}{|z|^2}\right).$$

We have $\Psi(|z|) = \Lambda |z|^{\alpha+2}$ for inverse-power laws, with $\alpha \geq -3$ and $\Lambda > 0$. The case $\alpha = -3$ is the so-called Coulombian case, of primary importance for applications. In such case the Boltzmann collision operator has no meaning, due to the divergence of the integral, even for smooth functions (a cut-off angular approximation is then used and the Landau equation can be derived in the so called grazing collision limit (Villani, 2002)).

Since conservation of mass, momentum, and energy, as well as H-theorem for the entropy are satisfied, equilibrium states are Maxwellians.

1.5.3 Additional models

- Enskog model: takes into account the nonlocality of the interactions induced by the diameter of the interacting spheres (accurately describes the behavior of dense gases). The collision operator is delocalized in space (regularization effect).
- Quantum-Boltzmann models: the nonlinear interactions $f'f'_* - ff_*$ is replaced by

  $$f'f'(1 \pm f)(1 \pm f_*) - ff_*(1 \pm f')(1 \pm f'_*).$$

  The minus sign corresponds corresponds to fermions (such as alectrons), and the plus sign to bosons (such as photons). The collision operator are called Pauli operator and Bose-Einstein operator respectively.
- Semiconductor-Boltzmann models: the linear Boltzmann equation for semiconductor devices has the form

  $$Q_S(f,M) = \int \sigma(v,v_*)\{M(v)f(v_*) - M(v_*)f(v)\}\, dv_*,$$

  where M is the normalized equilibrium distribution (Maxwellian, Fermi-Dirac) at the temperature θ of the lattice. The function $\sigma(v,v_*)$ describes the interaction of carriers with phonons.

- Granular gas models: particles undergo inelastic collisions. Energy is dissipated by the model and the steady states are Dirac delta function centered in the mean velocity.

More recently kinetic modelling has been applied to new fields as vehicular traffic flows, biomathematics (chemotaxis, inhalation of sprays, flocking), finance (modelling income distributions, price formation), coagulation-fragmentation processes, supply chains, and so on.

1.6 The splitting approach

The most common approach to solve numerically the full Boltzmann equation is based on an operator splitting (Desvillettes and Mischler, 1996).

The solution in one time step Δt may be obtained by the sequence of two steps. First integrate the space homogeneous equation for all $x \in \Omega$,

$$\begin{aligned} \frac{\partial \tilde{f}}{\partial t} &= \frac{1}{\varepsilon} Q(\tilde{f}, \tilde{f}), \\ \tilde{f}(x, v, 0) &= f_0(x, v), \end{aligned} \tag{32}$$

for a time step Δt (*collision step*) to obtain $\tilde{f} = \mathcal{C}_{\Delta t}(f_0)$, and then the transport equation using the output of the previous step as initial condition,

$$\begin{aligned} \frac{\partial f}{\partial t} + v \cdot \nabla_x f &= 0, \\ f(x, v, 0) &= \tilde{f}(x, v, \Delta t). \end{aligned} \tag{33}$$

for a time step Δt (*transport step*) to get $f = \mathcal{T}_{\Delta t}(\tilde{f}) = \mathcal{T}_{\Delta t}(\mathcal{C}_{\Delta t}(f_0))$.

After computing an approximation of the solution at time Δt, the process may be iterated to obtain the numerical solution at later times. Although this splitting scheme (simple splitting) described above is first order accurate in time it is very popular because it has several nice properties.

- The collision step acts only on v whereas the transport step acts on x. This makes the implementation of the resulting scheme simpler (it allows the use of any existing code designed to solve the free transport equation) and highly parallelizable.

- It makes simpler to design schemes which preserves the physical properties of the equation (conservations, positivity, H-theorem), since these properties essentially depends on the treatment of the collision step.

It is then clear, that after this splitting almost all the main numerical difficulties are contained in the collision step. The discretization of the resulting equations can be performed in a variety of ways (finite volume, finite difference, Monte Carlo methods and so on). The choice of the discretization mainly depends on the method that is used for the solution of the space homogeneous Boltzmann equation. Higher order splitting formulas can be derived in different ways (see Hairer et al. (2002)). For example the well-known second order Strang splitting (Strang, 1968) can be written as

$$\mathcal{C}_{\Delta t/2}(\mathcal{T}_{\Delta t}(\mathcal{C}_{\Delta t/2}(f_0))). \tag{34}$$

Unfortunately for splitting methods of order higher then two it can be shown that it's impossible to avoid negative time steps both in the transport as well as in the collision (Hairer et al., 2002). As an example we report here a symmetric fourth order formula (McLachlan, 1995)

$$\mathcal{T}_{a_1\Delta t}(\mathcal{C}_{b_1\Delta t}(\mathcal{T}_{a_2\Delta t}(\mathcal{C}_{b_2\Delta t}(\mathcal{T}_{a_3\Delta t}(\mathcal{C}_{b_2\Delta t}(\mathcal{T}_{a_2\Delta t}(\mathcal{C}_{b_1\Delta t}(\mathcal{T}_{a_1\Delta t}(f_0))))))))), \tag{35}$$

where

$$b_1 = \frac{6}{11}, \quad b_2 = \frac{1}{2} - b_1 \approx -0.045, \tag{36}$$

$$a_1 = \frac{642 + \sqrt{471}}{3924} \approx 0.169, \quad a_2 = \frac{121}{3924}(12 - \sqrt{471}) \approx -0.299, \quad a_3 = 1 - 2(a_1 + a_2). \tag{37}$$

Higher order formulas which avoid negative time stepping can be obtained as suitable combination of splitting steps (Dia and Schatzman, 1996). For example a third order approximation is given by

$$\frac{2}{3}[\mathcal{T}_{\Delta t/2}(\mathcal{C}_{\Delta t}(\mathcal{T}_{\Delta t/2}(f_0))) + \mathcal{C}_{\Delta t/2}(\mathcal{T}_{\Delta t}(\mathcal{C}_{\Delta t/2}(f_0)))] - \frac{1}{6}[\mathcal{T}_{\Delta t}(\mathcal{C}_{\Delta t}(f_0)) + \mathcal{C}_{\Delta t}(\mathcal{T}_{\Delta t}(f_0))], \tag{38}$$

which corresponds to take a combination of symmetrized Strang and first order splitting, whereas a fourth order scheme reads

$$\frac{4}{3}\mathcal{C}_{\Delta t/4}(\mathcal{T}_{\Delta t/2}(\mathcal{C}_{\Delta t/2}(\mathcal{T}_{\Delta t/2}(\mathcal{C}_{\Delta t/4}(f_0))))) - \frac{1}{3}\mathcal{C}_{\Delta t/2}(\mathcal{T}_{\Delta t}(\mathcal{C}_{\Delta t/2}(f_0))). \tag{39}$$

Clearly all the above splitting methods admit the symmetric formulation obtained by switching the transport and the collision operators. Note, however, that the appearance of negative coefficients or negative time steps in high order formulas may lead to some drawbacks in practical applications like the lack of positivity of the solution which makes very difficult their use in Monte Carlo schemes.

1.7 Asymptotic preserving methods

Even if it is difficult to give a rigorous definition of asymptotic preserving scheme since the concept has been used for a long time in the physics and mathematics literature and may refer to different discretization parameters (see Figure 3), here following Jin (1995) and Pareschi and Russo (2005), we formalize this notion for the time discretization of equation (1).

Definition 1 *A consistent time discretization method for (1) of stepsize* Δt *is* asymptotic preserving (AP) *if, independently of the initial data (2) and of the stepsize* Δt*, in the limit* $\varepsilon \to 0$ *becomes a consistent time discretization method for the reduced system (23).*

Note that this definition does not imply that the scheme preserves the order of accuracy in t in the stiff limit $\varepsilon \to 0$.

In the case of operator splitting we can reformulate the asymptotic preserving property and prove that

$$\begin{tikzcd}
P^{\varepsilon} \arrow[r, "\varepsilon \to 0"] & P^{0} \\
P_h^{\varepsilon} \arrow[u, "h \to 0"] \arrow[r, "\varepsilon \to 0"'] & P_h^{0} \arrow[u, "h \to 0"']
\end{tikzcd}$$

Figure 3: The AP diagram. Here P^{ε} is the original singular perturbation problem and P_h^{ε} its numerical approximation characterized by a discretization parameter h. The AP property corresponds to the request that P_h^{ε} is consistent with P^{ε} as $\varepsilon \to 0$ independently of h.

Proposition 1 *A sufficient condition for a consistent time discretization method of stepsize Δt applied to the operator splitting approximation of (1), given by (32)-(33), to be AP is that the time discretization of step (32), independently of the initial data (2) and of the stepsize Δt, in the limit $\varepsilon \to 0$ projects the solution f over the local Maxwellian equilibrium $M(\rho_0, u_0, T_0)$.*

The proof of the above proposition is an immediate consequence of the fact that as $\varepsilon \to 0$ step (32) degenerates into the projection $\mathcal{C}_{\Delta t}(f_0) = M(\rho_0, u_0, T_0)$ which coupled with the transport step (33) originates a so-called kinetic approximation (Coron and Perthame, 1991) to the Euler equation (23) given by $\mathcal{T}_{\Delta t}(M(\rho_0, u_0, T_0))$. We omit further details.

In other words, Proposition 1 states that if the relaxation step (32) is AP then the whole splitting (32)-(33) is AP. Analogous results hold true for the higher order splitting methods (34), (35), (38) and (39). Let us point out that degradation to first order accuracy when $\varepsilon \to 0$ is observed for most splitting methods like the one reviewed here. A possible way to overcome this drawback is based on the use of Implicit-Explicit (IMEX) Runge-Kutta methods for the full problem (1). We will discuss this aspect in Section 4.3.

For the sake of completeness we finally introduce the notion of entropic stability, namely schemes that preserve at a discrete level the entropy inequality (22). Let us denote by f^n, $n \geq 1$ the numerical solution at $t = n\Delta t$ obtained with a given time discretization method applied to (32) with initial data f_0.

Definition 2 *A time discretization method for (32) is called* unconditionally entropic *if $H(f^{n+1}) \leq H(f^n)$, where $H(f) = \int_{\mathbb{R}^3} f \log f \, dv$, independently of the step size Δt.*

As pointed out in Dimarco and Pareschi (2010a) and Higueras (2005), except for first order implicit Euler, the entropy inequality is not satisfied by high order implicit Runge-Kutta schemes applied to (32) unless a suitable time step restriction is considered. At variance exponential methods permits to construct unconditionally entropic methods at any order of accuracy (Gabetta et al., 1997).

In the next section we will focus on the solution to the space homogeneous Boltzmann equation (32). It is clear, in fact, that most computational challenges related to the behavior of the full equation depend on the way we approximate the collision operator.

2 Fast Boltzmann solvers

In this section we shall approximate the collision operator starting from a representation which somehow conserves more symmetries of the collision operator when one truncates it in a bounded domain. This representation was used in Bobylev and Rjasanow (1997), Bobylev and Rjasanow (1999), Bobylev and Rjasanow (2000), Ibragimov and Rjasanow (2002) and it's close to the classical Carleman representation (cf. Carleman (1932)). As we will see it is an essential step for the derivation of fast algorithms. The presentation here follows the line developed in Mouhot and Pareschi (2006), Mouhot and Pareschi (2004), Filbet et al. (2006), Mouhot and Pareschi (2011).

2.1 Restriction to bounded domains

The basic identity we shall need is

$$\frac{1}{2}\int_{\mathbb{S}^{d-1}} F(|u|\omega - u)\,d\omega = \frac{1}{|u|^{d-2}}\int_{\mathbb{R}^d} \delta(2\,x\cdot u + |x|^2)\,F(x)\,dx, \tag{40}$$

and can be verified easily by completing the square in the delta Dirac function, taking the spherical coordinate $x = r\,\omega$ and performing the change of variable $r^2 = s$.

Setting $u = v - v_*$ we can write the collision operator in the form

$$Q(f,f)(v) = \int_{v_*\in\mathbb{R}^d}\left\{\int_{\omega\in\mathbb{S}^{d-1}} B(\cos\theta,|u|)\right.$$
$$\left.\left[f\left(v_* - \frac{|u|\omega - u}{2}\right) f\left(v + \frac{|u|\omega - u}{2}\right) - f(v_*)\,f(v)\right]\,d\omega\right\} dv_*$$

and thus equation (40) yields

$$Q(f,f)(v) = 2\int_{v_*\in\mathbb{R}^d}\left\{\int_{x\in\mathbb{R}^d} B\left(\frac{x\cdot u}{|x||u|},|u|\right)\frac{1}{|u|^{d-2}}\,\delta(2\,x\cdot u + |x|^2)\right.$$
$$\left.\Big[f(v_* - x/2)\,f(v + x/2) - f(v_*)\,f(v)\Big]\,dx\right\} dv_*.$$

Now let us make the change of variable $x \to x/2$ in x to get

$$Q(f,f)(v) = 2^{d+1}\int_{v_*\in\mathbb{R}^d}\int_{x\in\mathbb{R}^d} B\left(\frac{x\cdot u}{|x||u|},|u|\right)\frac{1}{|u|^{d-2}}\,\delta(4\,x\cdot u + 4|x|^2)$$
$$[f(v_* - x)\,f(v + x) - f(v_*)\,f(v)]\,dx\,dv_*$$

and then setting $y = v_* - v - x$ in v_* we obtain

$$Q(f,f)(v) = 2^{d+1} \int_{y\in\mathbb{R}^d}\int_{x\in\mathbb{R}^d} B\left(\frac{x\cdot u}{|x||u|}, |u|\right) \frac{1}{|u|^{d-2}}\,\delta(-4x\cdot y)$$
$$[f(v+y)\,f(v+x) - f(v+x+y)\,f(v)]\;dx\,dy$$

where now $u = -(x+y)$. Thus in the end we have

$$Q(f,f)(v) = 2^{d-1} \int_{x\in\mathbb{R}^d}\int_{y\in\mathbb{R}^d} B\left(-\frac{x\cdot(x+y)}{|x||x+y|}, |x+y|\right) \frac{1}{|x+y|^{d-2}}\,\delta(x\cdot y)$$
$$[f(v+y)\,f(v+x) - f(v+x+y)\,f(v)]\;dx\,dy.$$

Figure 4 sums up the different geometrical quantities of the usual representation and the one we derived from Carleman's one.

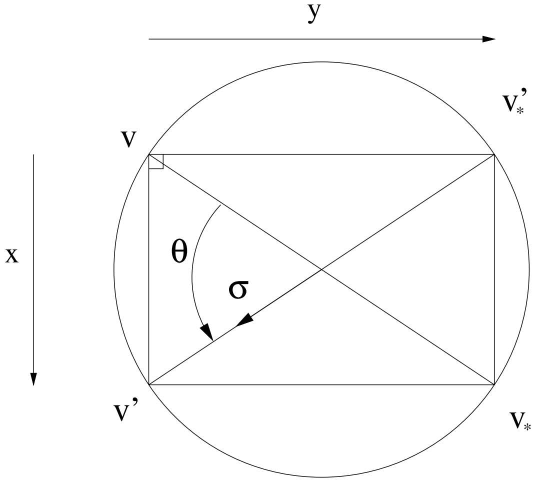


Figure 4: Geometry of the collision $(v, v_*) \leftrightarrow (v', v'_*)$.

Now let us consider the bounded domain $\mathcal{D}_T = [-T, T]^d$ $(0 < T < +\infty)$. There are two possibilities of truncation to reduce the collision process in a box. From now on let us write

$$\tilde{B}(x,y) = 2^{d-1}\,B\left(-\frac{x\cdot(x+y)}{|x||x+y|}, |x+y|\right)\,|x+y|^{-(d-2)}.$$

One can easily see that on the manifold defined by $x\cdot y = 0$, a simpler formula is

$$\tilde{B}(x,y) = \tilde{B}(|x|,|y|) = 2^{d-1}\,B\left(\frac{|x|}{\sqrt{|x|^2+|y|^2}}, \sqrt{|x|^2+|y|^2}\right)\,(|x|^2+|y|^2)^{-\frac{d-2}{2}}. \qquad (41)$$

First one can remove some physical collisions connecting with some points out of the box. This is the natural preliminary stage for deriving conservative schemes. In this case there is no need for a truncation on the modulus of x and y since we impose them to stay in the box. It yields

$$Q^{\mathrm{tr}}(f,f)(v) = \int\!\!\int_{\left\{x,\,y\,\in\,\mathbb{R}^d\;|\;v+x,\,v+y,\,v+x+y\,\in\,\mathcal{D}_T\right\}} \tilde{B}(x,y)\,\delta(x\cdot y)$$
$$[f(v+y)\,f(v+x) - f(v+x+y)\,f(v)]\;dx\,dy$$

defined for $v \in \mathcal{D}_T$. One can easily check that the following weak form is satisfied by this operator

$$\int Q^{\mathrm{tr}}(f,f)\,\varphi(v)\,dv = \frac{1}{4}\int\int\int_{\left\{v,\,x,\,y\,\in\,\mathbb{R}^d\ \mid\ v,\,v+x,\,v+y,\,v+x+y\,\in\,\mathcal{D}_T\right\}} \tilde{B}(x,y)\,\delta(x\cdot y)$$
$$f(v+x+y)\,f(v)\,[\varphi(v+y)+\varphi(v+x)-\varphi(v+x+y)-\varphi(v)]\;dv\,dx\,dy \quad (42)$$

and this implies conservation of mass, momentum and energy as well as the H-theorem on the entropy. Note that at this level this formulation gives no advantage with respect to the usual one obtained from (3) by restricting $v, v_*, v', v'_* \in \mathcal{D}_T$. The problem of this truncation is that it corresponds to change the collision kernel by adding some artificial dependence on v, v_*, v', v'_*. In this way convolution-like properties are broken.

A different approach consists in periodizing the function f on the domain $\mathcal{D}_T$. Here we have to truncate the integration in x and y since periodization would yield infinite result if not. Thus we set them to vary in $\mathcal{B}_R$, the ball of center 0 and radius R. Then a geometrical argument (see Pareschi and Russo (2000b)) shows that using the periodicity of the function it is enough to take $T \geq (3+\sqrt{2})R/2$ to prevent intersections of the regions where f is different from zero.

The operator now reads

$$Q^R(f,f)(v) = \int_{x\in\mathcal{B}_R}\int_{y\in\mathcal{B}_R} \tilde{B}(x,y)\,\delta(x\cdot y)$$
$$[f(v+y)f(v+x)-f(v+x+y)f(v)]\;dx\,dy \quad (43)$$

for $v \in \mathcal{D}_T$ (the expression for $v \in \mathbb{R}^d$ is deduced by periodization). The interest of this representation is to preserve the real collision kernel and its properties.

By making some translation changes of variable on v (by x, y and $x+y$), using the changes $x \to -x$ and $y \to -y$ and the fact that

$$\tilde{B}(-x,y)\,\delta(-x\cdot y) = \tilde{B}(x,y)\,\delta(x\cdot y) = \tilde{B}(x,-y)\,\delta(x\cdot -y)$$

one can easily prove that for any function φ *periodic* on $\mathcal{D}_T$ the following weak form is satisfied

$$\int_{\mathcal{D}_T} Q^R(f,f)\,\varphi(v)\,dv = \frac{1}{4}\int_{v\in\mathcal{D}_T}\int_{x\in\mathcal{B}_R}\int_{y\in\mathcal{B}_R} \tilde{B}(x,y)\,\delta(x\cdot y)$$
$$f(v+x+y)f(v)\,[\varphi(v+y)+\varphi(v+x)-\varphi(v+x+y)-\varphi(v)]\;dv\,dx\,dy. \quad (44)$$

About the conservation properties one can shows that if f has compact support included in $\mathcal{B}_R$ with $T \geq (3+\sqrt{2})R/2$ (no aliasing condition, see Pareschi and Russo (2000b) for a detailed discussion), then no unphysical collisions occur and thus mass, momentum and energy are preserved. Obviously this compactness is not preserved with time since the collision operator spreads the support of f by a factor $\sqrt{2}$. In the rest of the paper we will focus on the periodized truncation Q^R.

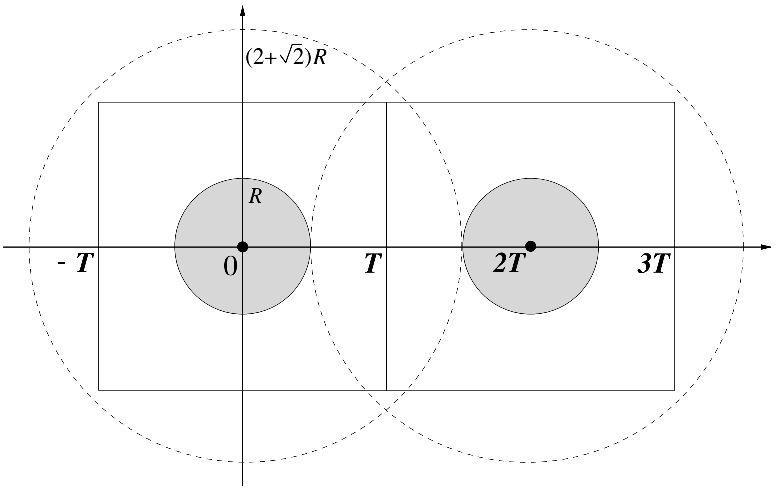


Figure 5: Periodization in 2D

2.2 Spectral methods

Now we use the representation Q^R to derive new spectral methods. The spectral methods for kinetic equations originated in the works of Pareschi and Perthame (1996), Pareschi and Russo (2000b), and were further developed in Pareschi and Russo (2000c) and Filbet and Russo (2003). Before they had a long history in fluid mechanics, see Canuto et al. (1988).

To simplify notations let us take $T = \pi$. Hereafter we use just one index to denote the d-dimensional sums of integers.

The approximate function f_N is represented as the truncated Fourier series

$$f_N(v) = \sum_{k=-N}^{N} \hat{f}_k e^{ik\cdot v},$$

$$\hat{f}_k = \frac{1}{(2\pi)^d} \int_{\mathcal{D}_\pi} f(v) e^{-ik\cdot v}\, dv.$$

The spectral equation is the projection of the collision equation in $\mathbb{P}^N$, the $(2N+1)^d$-dimensional vector space of trigonometric polynomials of degree at most N in each direction, i.e

$$\frac{\partial f_N}{\partial t} = \mathcal{P}_N Q^R(f_N, f_N)$$

where $\mathcal{P}_N$ denotes the orthogonal projection on $\mathbb{P}^N$ in $L^2(\mathcal{D}_\pi)$. A straightforward computation leads to the following set of ordinary differential equations on the Fourier coefficients

$$\hat{f}'_k(t) = \sum_{\substack{l,m=-N \\ l+m=k}}^{N} \hat{\beta}(l,m)\, \hat{f}_l\, \hat{f}_m, \quad k = -N, ..., N \tag{45}$$

where $\hat{\beta}(l,m)$ are the so-called *kernel modes*, given by

$$\hat{\beta}(l,m) = \int_{x\in\mathcal{B}_R} \int_{y\in\mathcal{B}_R} \tilde{B}(x,y)\, \delta(x\cdot y)\, \left[e^{il\cdot x}\, e^{im\cdot y} - e^{im\cdot(x+y)}\right] dx\, dy.$$

The kernel modes can be written as

$$\hat{\beta}(l,m) = \beta(l,m) - \beta(m,m)$$

where

$$\beta(l,m) = \int_{x\in\mathcal{B}_R}\int_{y\in\mathcal{B}_R} \tilde{B}(x,y)\,\delta(x\cdot y)\,e^{il\cdot x}\,e^{im\cdot y}\,dx\,dy.$$

Therefore in the sequel we shall focus on β, and one easily checks that $\beta(l,m)$ depends only on $|l|$, $|m|$ and $|l\cdot m|$.

Finally let us compare the new kernel modes with the ones in Pareschi and Russo (2000b). The usual kernel modes written in the x and y variables reads

$$\hat{\beta}_{\text{usual}}(l,m) = \int_{x\in\mathcal{B}_R}\int_{y\in\mathcal{B}_R} \tilde{B}(x,y)\,\delta(x\cdot y)\,\chi_{\{|x+y|\le R\}}\,\left[e^{il\cdot x}\,e^{im\cdot y} - e^{im\cdot(x+y)}\right]\,dx\,dy.$$

Thus the usual representation contains a strong coupling between x and y which makes it very hard the construction of fast algorithms.

2.3 Discrete-velocity models

The representation Q^R of this section can also be used to derive fast solvers for discrete velocity models (DVM). Historically these methods were among the first deterministic methods for discretizing the Boltzmann equation in velocity space. The discretization is built starting from physical rather then numerical considerations. We assume the gas particles can attain only a finite set of velocities

$$V_N = \{v_1, v_2, v_3, \ldots, v_N\}, \quad v_i \in I\!R^3.$$

Any DVM can be written as a product quadrature formula for (3) in the general form

$$D_i = \sum_{j,k,l\,\in\,\mathbb{Z}^d} \Gamma_{i,j}^{k,l}\left[f_k f_l - f_i f_j\right],$$

where D_i denotes the discrete Boltzmann collision operator and the integer indexes refer to the points in the computational grid. In order to keep conservations the coefficients $\Gamma_{i,j}^{k,l}$ are defined by

$$\Gamma_{i,j}^{k,l} = \mathbf{1}(i+j-k-l)\,\mathbf{1}(|i|^2+|j|^2-|k|^2-|l|^2)\,B(|k-i|,|l-j|)\,w_{i,j}^{k,l}$$

where $\mathbf{1}$ denotes the function on $\mathbb{Z}$ defined by $\mathbf{1}(z)=1$ if $z=0$ and 0 elsewhere, and $w_{i,j}^{k,l}>0$ are the weights of the quadrature formula, which characterize the different DVM. $B>0$ is the discrete collision kernel. One can check on this formulation that the scheme satisfies the usual conservation laws and entropy inequality (see Platkowski and Illner (1988) and the references therein).

We can write at the discrete level the same representation as in the continuous case

$$D_i = \sum_{k,l\,\in\,\mathbb{Z}^d} \tilde{\Gamma}_{k,l}\left[f_{i+k}f_{i+l} - f_i f_{i+k+l}\right]$$

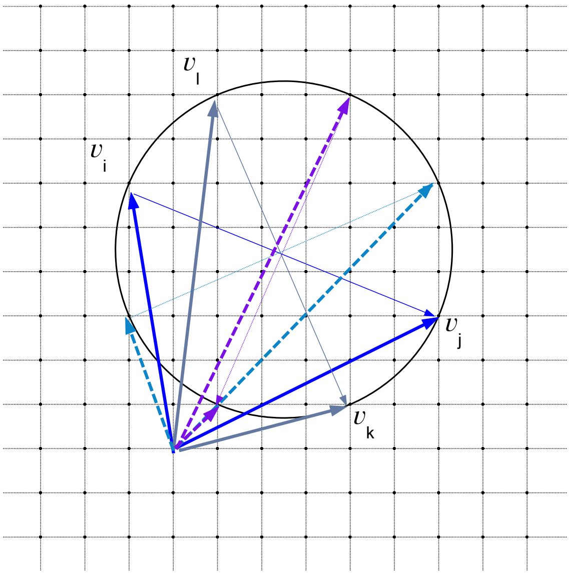


Figure 6: Sketch of a planar model based on a cartesian grid. Note that in general few grid points will belong to the collision circle.

with

$$\tilde{\Gamma}_{k,l} = \mathbf{1}(k \cdot l)\, \frac{B(|k|,|l|)}{|k+l|}\, w_{k,l}.$$

This is coherent with the DVM obtained by quadrature starting from the Carleman representation in Panferov and Heintz (2002).

Now again when one is interested to compute the DVM in a bounded domain there are two possibilities. First as in the case of $Q^{\rm tr}$ one can force the discrete velocities to stay in a box, which yields for $i = -N, \dots, N$ (again using the one index notation for d-dimensional sums)

$$D_i^{\rm tr} = \sum_{\substack{k,l \\ -N \le i+k,\, i+l,\, i+k+l \le N}} \tilde{\Gamma}_{k,l} \big[f_{i+k} f_{i+l} - f_i f_{i+k+l} \big].$$

This new discrete operator is completely conservative but the collision kernel is not invariant anymore according to i, which breaks the convolution properties.

The other possibility is to periodize the function f over the box and truncate the sum in k and l. It yields for a given truncation parameter $\tilde{N} \in \mathbb{N}$

$$D_i^{\tilde{N}} = \sum_{-\tilde{N} \le k,l \le \tilde{N}} \tilde{\Gamma}_{k,l} \big[f_{i+k} f_{i+l} - f_i f_{i+k+l} \big], \qquad (46)$$

for any $i = -N \dots N$.

It is easy to see that $D^{\tilde{N}}$ satisfies exactly a discrete weak form and conservation properties similar to Q^R. Moreover one can derive the following consistency result from (Panferov and Heintz, 2002, Theorem 3) in the case of hard spheres collision kernel

Theorem 1 *Assume that $f, g \in C^k(\mathbb{R}^3)$ $(k \geq 1)$ with compact support $\mathcal{B}_S$. The uniform grid of step h is constructed on the box $\mathcal{D}_T$ with the no aliasing condition $T \geq (3+\sqrt{2})S/2$. Then for $\tilde{N} = [S/h]$ (where $[\,\cdot\,]$ denotes the integer value) and $h > 0$ sufficiently small,*

$$\|Q(g,f) - D_h^{\tilde{N}}(g,f)\|_{L^\infty(\mathbb{Z}_h)} \leq C\, h^r$$

where $D_h^{\tilde{N}}$ is the DVM operator defined in (46) (for the precise quadrature weights derived in Panferov and Heintz (2002)) on the grid above-mentioned, and $f_i = f(ih)$. Here $r = k/(k+3)$ and the constant C is independent on h.

Remark 2 *As can be seen from Theorem 1, the periodized DVM presented in this subsection is expected to have a quite low accuracy. On the contrary the spectral method will be proven to be spectrally accurate, i.e. of infinite order for smooth solutions. Nevertheless this periodized DVM has some interesting features compared to the spectral method. Indeed, one can prove that if the quadrature weights $w_{k,l}$ are non-negative, then the scheme is stable in the sense that if one starts from a non-negative initial data, then the solution remains non-negative and has thus constant L^1 norm. Concerning the spectral method we refer to Filbet and Mouhot (2011) for an analysis of its stability and convergence properties.*

2.4 Fast spectral algorithms

As soon as one is searching for fast deterministic algorithms for the collision operator, i.e algorithm with a cost lower than $O(N^{2d+\varepsilon})$ (which is the cost of a usual discrete velocity model, with typically $\varepsilon = 1$), one has to find some way to compute the collision operator *without going through all the couples of collision points* during the computation. This leads naturally to search for some convolution structure (discrete or continuous) in the operator. Unfortunately, as discussed in the previous sections, this is rather contradictory with the search for a conservative scheme in a bounded domain, since the boundary condition needed to prevent for the outgoing or ingoing collisions breaks the invariance.

Here we search for a convolution structure in the equations (45). The aim is to approximate each $\hat{\beta}(l,m)$ by a sum

$$\hat{\beta}(l,m) \simeq \sum_{p=1}^{A} \alpha_p(l)\alpha_p'(m).$$

This gives a sum of A discrete convolutions and so the algorithm can be computed in $O(A\, N^d \log_2 N)$ operations by means of standard FFT techniques (Canuto et al., 1988; Cooley and Tukey, 1965). Obviously this is equivalent to obtain such a decomposition on β. To this purpose we shall use a further approximated collision operator where the number of possible directions of collision is reduced to a finite set.

2.4.1 A semi-discrete collision operator

We write x and y in spherical coordinates

$$Q^R(f,f)(v) = \frac{1}{4}\int_{e\in\mathbb{S}^{d-1}}\int_{e'\in\mathbb{S}^{d-1}} \delta(e\cdot e')\, de\, de'$$
$$\left\{\int_{-R}^{R}\int_{-R}^{R} \rho^{d-2}\,(\rho')^{d-2}\,\tilde{B}(\rho,\rho')\,[f(v+\rho'e')f(v+\rho e) - f(v+\rho e+\rho'e')f(v)]\; d\rho\, d\rho'\right\}. \tag{47}$$

Let us take $\mathcal{A}$ a discrete set of orthogonal couples of unit vectors (e,e'), which is even: $(e,e')\in\mathcal{A}$ implies that $(-e,e')$, $(e,-e')$ and $(-e,-e')$ belong to $\mathcal{A}$ (this property on the set $\mathcal{A}$ is required to preserve the conservation properties of the operator). Now we define $Q_R^{\mathcal{A}}$ to be

$$Q^{R,\mathcal{A}}(f,f)(v) = \frac{1}{4}\int_{(e,e')\in\mathcal{A}} \left\{\int_{-R}^{R}\int_{-R}^{R} \rho^{d-2}\,(\rho')^{d-2}\,\tilde{B}(\rho,\rho')\big[f(v+\rho'e')f(v+\rho e) - \right.$$
$$\left. f(v+\rho e+\rho'e')f(v)\big]\, d\rho\, d\rho'\right\} d\mathcal{A}$$

where $d\mathcal{A}$ denotes a discrete measure on $\mathcal{A}$ which is also even in the sense that $d\mathcal{A}(e,e') = d\mathcal{A}(-e,e') = d\mathcal{A}(e,-e') = d\mathcal{A}(-e,-e')$. Using again translation change of variable on v by ρe, $\rho'e'$ and $\rho e + \rho'e'$ and the symmetries of the set $\mathcal{A}$ one can easily derive the following weak form on $Q_R^{\mathcal{A}}$. For any function φ *periodic* on $\mathcal{D}_T$,

$$\int_{\mathcal{D}_T} Q^{R,\mathcal{A}}(f,f)\,\varphi(v)\, dv = \frac{1}{16}\int_{v\in\mathcal{D}_T}\int_{(e,e')\in\mathcal{A}}\int_{-R}^{R}\int_{-R}^{R} \rho^{d-2}\,(\rho')^{d-2}\,\tilde{B}(\rho,\rho')$$
$$f(v+\rho e+\rho'e')f(v)\,[\varphi(v+\rho'e') + \varphi(v+\rho e) - \varphi(v+\rho e+\rho'e') - \varphi(v)]\; d\rho\, d\rho'\, d\mathcal{A}\, dv.$$

This immediately gives the same conservations properties as Q_R.

2.4.2 Expansion of the kernel modes

We make the *decoupling assumption* that

$$\tilde{B}(x,y) = a(|x|)\, b(|y|). \tag{48}$$

This assumption is obviously satisfied if $\tilde{B}$ is constant. This is the case of Maxwellian molecules in dimension two, and hard spheres in dimension three (the most relevant kernel for applications). Extensions to more general interactions are discussed in Mouhot and Pareschi (2006).

First let us deal with dimension 2 with $\tilde{B} = 1$ to explain the method. Here we write x and y in spherical coordinates $x = \rho e$ and $y = \rho'e'$ to get

$$\beta(l,m) = \frac{1}{4}\int_{e\in\mathbb{S}^1}\int_{e'\in\mathbb{S}^1} \delta(e\cdot e')\left[\int_{-R}^{R} e^{i\rho(l\cdot e)}\, d\rho\right]\left[\int_{-R}^{R} e^{i\rho'(m\cdot e')}\, d\rho'\right] de\, de'.$$

Let us denote by

$$\phi^2_R(s) = \int_{-R}^{R} e^{i\rho s}\, d\rho,$$

for $s \in \mathbb{R}$. It is easy to see that ϕ^2_R is even and we can give the explicit formula

$$\phi^2_R(s) = 2\, R\, \mathrm{Sinc}(Rs)$$

with $\mathrm{Sinc}(\theta) = (\sin\theta)/\theta$.

Thus we have

$$\beta(l,m) = \frac{1}{4} \int_{e\in\mathbb{S}^1} \int_{e'\in\mathbb{S}^1} \delta(e\cdot e')\, \phi^2_R(l\cdot e)\, \phi^2_R(m\cdot e')\, de\, de'$$

and thanks to the parity property of ϕ^2_R we can adopt the following periodic parametrization

$$\beta(l,m) = \int_0^{\pi} \phi^2_R(l\cdot e_\theta)\, \phi^2_R(m\cdot e_{\theta+\pi/2})\, d\theta.$$

The function $\theta \to \phi^2_R(l\cdot e_\theta)\, \phi^2_R(m\cdot e_{\theta+\pi/2})$ is periodic on $[0,\pi]$ and thus the rectangular quadrature rule is of infinite order and optimal. A regular discretization of M equally spaced points thus gives

$$\beta(l,m) = \frac{\pi}{M} \sum_{p=0}^{M-1} \alpha_p(l)\alpha'_p(m)$$

with

$$\alpha_p(l) = \phi^2_R(l\cdot e_{\theta_p}), \qquad \alpha'_p(m) = \phi^2_R(m\cdot e_{\theta_p+\pi/2})$$

where $\theta_p = \pi p/M$.

More generally under the decoupling assumption (48) on $\tilde{B}$, we get the following decomposition formula

$$\beta(l,m) = \frac{\pi}{M} \sum_{p=0}^{M-1} \alpha_p(l)\alpha'_p(m)$$

where

$$\alpha_p(l) = \phi^2_{R,a}(l\cdot e_{\theta_p}), \qquad \alpha'_p(m) = \phi^2_{R,b}(m\cdot e_{\theta_p+\pi/2})$$

and

$$\phi^2_{R,a}(s) = \int_{-R}^{R} a(\rho)\, e^{i\rho s}\, d\rho, \qquad \phi^2_{R,b}(s) = \int_{-R}^{R} b(\rho')\, e^{i\rho' s}\, d\rho'$$

with $\theta_p = \pi p/M$.

Remark 3 *In the symmetric case $a = b$ (for instance for hard spheres) it is possible to parametrize $\beta(l,m)$ as*

$$\beta(l,m) = 2 \int_0^{\pi/2} \phi^2_{R,a}(l\cdot e_\theta)\, \phi^2_{R,a}(m\cdot e_{\theta+\pi/2})\, d\theta$$

and the function $\theta \to \phi^2_{R,a}(l\cdot e_\theta)\, \phi^2_{R,a}(m\cdot e_{\theta+\pi/2})$ is periodic on $[0,\pi/2]$. Thus the decomposition can be obtained by applying the rectangular rule on this interval. At the numerical level it yields a reduction of the cost by a factor 2.

Now let us deal with dimension $d = 3$ with $\tilde{B}$ satisfying the decoupling assumption (48). First we change to the spherical coordinates

$$\beta(l,m) = \frac{1}{4}\int_{e\in\mathbb{S}^2}\int_{e'\in\mathbb{S}^2}\delta(e\cdot e')\left[\int_{-R}^{R}\rho\, a(\rho)\, e^{i\rho(l\cdot e)}\, d\rho\right]\left[\int_{-R}^{R}\rho'\, b(\rho')\, e^{i\rho'(m\cdot e')}\, d\rho'\right] de\, de'$$

and then we integrate first e' on the intersection of the unit sphere with the plane $e^\perp$,

$$\beta(l,m) = \frac{1}{4}\int_{e\in\mathbb{S}^2}\phi^3_{R,a}(l\cdot e)\left[\int_{e'\in\mathbb{S}^2\cap e^\perp}\phi^3_{R,b}(m\cdot e')\, de'\right] de$$

where

$$\phi^3_{R,a}(s) = \int_{-R}^{R}\rho\, a(\rho)\, e^{i\rho s}\, d\rho.$$

Thus we get the following decoupling formula with two degrees of freedom

$$\beta(l,m) = \int_{e\in\mathbb{S}^2_+}\phi^3_{R,a}(l\cdot e)\,\psi^3_{R,b}\big(\Pi_{e^\perp}(m)\big)\, de$$

where $\mathbb{S}^2_+$ denotes the half-sphere and

$$\psi^3_{R,b}\big(\Pi_{e^\perp}(m)\big) = \int_0^\pi \sin\theta\,\phi_{R,b}\big(|\Pi_{e^\perp}(m)|\,\cos\theta\big)\, d\theta,$$

(this formula can be derived performing the change of variable $de' = \sin\theta\, d\theta\, d\varphi$ with the basis $(e, u = \Pi_{e^\perp}(m)/|\Pi_{e^\perp}(m)|, e\times u)$).

Again in the particular case where $\tilde{B} = 1$ (hard spheres model), we can compute explicitly the functions ϕ^3_R (in this case $a = b = 1$),

$$\phi^3_R(s) = R^2\,\left[2\mathrm{Sinc}(Rs) - \mathrm{Sinc}^2(Rs/2)\right], \qquad \psi^3_R(s) = 2\, R^2\,\mathrm{Sinc}^2(Rs/2).$$

Now the function $e \to \phi^3_{R,a}(l\cdot e)\,\psi^3_{R,b}\big(\Pi_{e^\perp}(m)\big)$ is periodic on $\mathbb{S}^2_+$ and so the rectangular rule is of infinite order and optimal. Taking a spherical parametrization (θ,φ) of $e \in \mathbb{S}^2_+$ and uniform grids of respective size M_1 and M_2 for θ and φ we get

$$\beta(l,m) = \frac{\pi^2}{M_1 M_2}\sum_{p,q=0}^{M_1,M_2}\alpha_{p,q}(l)\alpha'_{p,q}(m)$$

where

$$\alpha_{p,q}(l) = \phi^3_{R,a}(l\cdot e_{(\theta_p,\varphi_q)}), \qquad \alpha'_{p,q}(m) = \psi^3_{R,b}(\Pi_{e^\perp_{(\theta_p,\varphi_q)}}(m))$$

and

$$\phi^3_{R,a}(s) = \int_{-R}^{R}\rho\, a(\rho)\, e^{i\rho s}\, d\rho, \qquad \psi^3_{R,b}(s) = \int_0^\pi \sin\theta\,\phi^3_{R,b}(s\cos\theta)\, d\theta$$

and

$$(\theta_p,\varphi_q) = \Big(\frac{p\,\pi}{M_1}, \frac{q\,\pi}{M_2}\Big).$$

From now on we shall consider this expansion with $M = M_1 = M_2$ to avoid anisotropy in the computational grid.

Remark 4

For any dimension, we can construct as above an approximated collision operator $Q^{R,\mathcal{A}_M}$ *with*

$$\mathcal{A}_M = \left\{(e,e') \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1} \mid e \in \mathbb{S}^{d-1}_{M,+}, \quad e' \in e^\perp \cap \mathbb{S}^{d-1}\right\}$$

where $\mathbb{S}^{d-1}_{M,+}$ *denotes a uniform angular discretization of the half sphere with* M *points in each angular coordinate (the other half sphere is obtained by parity). Let us remark that this discretization contains exactly* M^{d-1} *points. From now on we shall denote*

$$Q^{R,M} = Q^{R,\mathcal{A}_M} = \sum_{p=1}^{M^{d-1}} Q_p^{R,M}.$$

2.4.3 Spectral accuracy

In this paragraph we are interested in computing the accuracy of the scheme according to the three parameters N (the number of modes), R (the truncation parameter), and M (the number of angular directions for each angular coordinate). Instead of looking at the error on each kernel mode it is more convenient to look at the error on the global operator. Here the Lebesgue spaces L^p, $p = 1 \ldots + \infty$, and the periodic Sobolev spaces H_p^k, $k = 0 \ldots + \infty$ refer to $\mathcal{D}_\pi$.

So in order to give a consistency result, the first step will be to prove a consistency result for the approximation of Q^R by $Q^{R,M}$ (see Mouhot and Pareschi (2006) for details).

Lemma 1 *The error on the approximation of the collision operator is spectrally small, i.e for all* $k > d-1$ *such that* $f \in H_p^k$

$$\|Q^R(g,f) - Q^{R,M}(g,f)\|_{L^2} \leq C_1 \, \frac{R^k \|g\|_{H_p^k} \|f\|_{H_p^k}}{M^k}.$$

For the second step we shall use the consistency result (Pareschi and Russo, 2000b, Corollary 5.4) on the operator Q^R, which we quote here for the sake of clarity.

Lemma 2 *For all* $k \in \mathbb{N}$ *such that* $f \in H_p^k$

$$\|Q^R(f,f) - \mathcal{P}_N Q^R(f_N,f_N)\|_{L^2} \leq \frac{C_2}{N^k}\left(\|f\|_{H_p^k} + \|Q^R(f_N,f_N)\|_{H_p^k}\right).$$

Combining these two results, one gets the following consistency result

Theorem 2 *For all* $k > d-1$ *such that* $f \in H_p^k(\mathcal{D}_\pi)$

$$\|Q^R(f,f) - \mathcal{P}_N Q^{R,M}(f_N,f_N)\|_{L^2} \leq C_1 \, \frac{R^k \|f_N\|^2_{H_p^k}}{M^k} + \frac{C_2}{N^k}\left(\|f\|_{H_p^k} + \|Q^R(f_N,f_N)\|_{H_p^k}\right).$$

Now let us focus briefly on the macroscopic quantities. First with Lemma 1 at hand one can establish the estimate

$$\|Q^{R,M}(g,f)\|_{L^2} \leq C \, \|g\|_{H_p^d} \|f\|_{H_p^d},$$

for a constant uniform in M. Then following the method of (Pareschi and Russo, 2000b, Remark 5.4) and using this estimate we obtain the following spectral accuracy result

$$\left|\langle Q^{R,M}(f,f),\varphi\rangle - \langle \mathcal{P}_N Q^{R,M}(f_N,f_N),\varphi\rangle\right|_{L^2} \leq \frac{C_3}{N^k}\,\|\varphi\|_{L^2}\left(\|f\|_{H_p^{k+d}} + \|Q^{R,M}(f_N,f_N)\|_{H_p^k}\right)$$

where φ can be replaced by $v, |v|^2$. Indeed there is no need to compare the momenta of $\mathcal{P}_N Q^{R,M}(f_N,f_N)$ with those of $Q^R(f,f)$ since $Q^{R,M}$ is also conservative, and so they can be compared directly to those of $Q^{R,M}$. Thus the error on momentum and energy is independent on M and is spectrally small according to N even for very small value of the parameter M.

2.4.4 Implementation aspects

The final spectral scheme depends on the three parameters N, R, and M. The only conditions on these parameters is the no-aliasing condition that relates R and the size of the box T (here π). A detailed study of the influence of the choices of N and R has been done in Pareschi and Russo (2000b). Here we are interested only in the influence of M over the computations, since M controls the computations speed-up.

The method of the previous subsections yields a decomposition of the collision operator, which after projection on $\mathbb{P}^N$ gives the following decomposition

$$\mathcal{P}_N Q^{R,M} = \sum_{p=1}^{M^{d-1}} \mathcal{P}_N Q_p^{R,M}. \tag{49}$$

Each $\mathcal{P}_N Q_p^{R,M}$ can be computed with a cost $O(N^d \log_2 N)$. Thus for a general choice of M and N we obtain the cost $O(M^{d-1} N^d \log_2 N)$. The decomposition (49) is completely parallelizable and thus the cost can be strongly reduced on a parallel machine (formally up to $O(N^d \log_2 N)$). One just has to make independent computations for the M^{d-1} terms of the decomposition.

Moreover the formula of decomposition is naturally adaptive (that is the number M can be made space dependent), which can be quite useful in the inhomogeneous setting, where some regions deserve less accuracy than others. Since it relies on the rectangular formula, whose adaptivity property is well known, one can easily double the number of directions M if needed, without computing again those points already computed.

Finally the decomposition can be also interesting from the storage viewpoint, as the classical spectral method requires the storage of a $N^d \times N^d$ matrix whereas the fast method requires the storage of $2M^{d-1}$ vectors of size N^d. In dimension 2 the classical method requires a storage of order $O(N^4)$ and the fast method requires a storage of order $O(MN^2)$. In dimension 3 the classical method requires a storage of order $O(N^4)$ (thanks to the symmetries of the matrix of kernel modes, see Pareschi and Russo (2000b)), and the fast method requires a storage of order $O(M^2\,N^3)$.

2.5 Fast DVM's algorithms

The fast algorithms developed for the spectral method can be in fact extended to the periodized DVM method. The method that originates is in some sense related to the direct FFT approach proposed in Bobylev and Rjasanow (1997, 2000, 1999).

2.5.1 Principle of the method: a pseudo-spectral viewpoint

We start from the periodized DVM in $[|-N,N|]^d$ with representation (46) and as in the continuous case we set, for $-\tilde{N} \le k, l \le \tilde{N}$,

$$\tilde{B}(|k|,|l|) = \frac{B(|k|,|l|)}{|k+l|^{d-2}} = 2^{d-1}\, B\left(\frac{|k|}{\sqrt{|k|^2+|l|^2}}, \sqrt{|k|^2+|l|^2}\right)\, (|k|^2+|l|^2)^{-\frac{d-2}{2}}.$$

With this notation

$$\tilde{\Gamma}_{k,l} = \mathbf{1}(k\cdot l)\, \tilde{B}(|k|,|l|)\, w_{k,l},$$

and thus the DVM becomes

$$f_i' = \sum_{-\tilde{N}\le k,l\le \tilde{N}} \mathbf{1}(k\cdot l)\, \tilde{B}(|k|,|l|)\, w_{k,l}\left[f_{i+k}f_{i+l} - f_i f_{i+k+l}\right].$$

Now we transform this set of ordinary differential equations into a new one using the involution transformation of the discrete Fourier transform on the vector $(f_i)_{-N\le i\le N}$. This involution reads

$$\tilde{f}_I = \frac{1}{2N+1}\sum_{i=0}^{2N} f_i\, \mathbf{e}_{-I}(i), \qquad f_i = \sum_{I=-N}^{N} \tilde{f}_I\, e_I(i)$$

where $\mathbf{e}_K(k)$ denotes $e^{\frac{2i\pi\, K\cdot k}{2N+1}}$, and thus the set of differential equations becomes

$$\begin{aligned}\tilde{f}_I' \;&=\; \sum_{K,L=-N}^{N}\left(\frac{1}{2N+1}\sum_{i=0}^{2N}\mathbf{e}_{K+L-I}(i)\right)\\ &\left[\sum_{-\tilde{N}\le k,l\le\tilde{N}} \mathbf{1}_{(k\cdot l)}\, \tilde{B}(|k|,|l|)\, w_{k,l}\Big(\mathbf{e}_K(k)\mathbf{e}_L(l) - \mathbf{e}_L(k+l)\Big)\right]\tilde{f}_K\,\tilde{f}_L\end{aligned}$$

for $-N\le I\le N$. We have the following identity

$$\frac{1}{2N+1}\sum_{i=0}^{2N}\mathbf{e}_{K+L-I}(i) = \mathbf{1}(K+L-I)$$

and so the set of equations is

$$\tilde{f}_I' = \sum_{\substack{K+L=I\\ K,L=-N}}^{N} \tilde{\beta}(K,L)\,\tilde{f}_K\,\tilde{f}_L$$

with

$$\tilde{\beta}(K,L) = \sum_{-\tilde{N}\le k,l\le\tilde{N}} \mathbf{1}(k\cdot l)\,\tilde{B}(|k|,|l|)\, w_{k,l}\left[\mathbf{e}_K(k)\mathbf{e}_L(l) - \mathbf{e}_L(k+l)\right] = \beta(K,L) - \beta(L,L)$$

where

$$\beta(K,L) = \sum_{-\tilde{N}\le k,l\le\tilde{N}} \mathbf{1}(k\cdot l)\,\tilde{B}(|k|,|l|)\, w_{k,l}\,\mathbf{e}_K(k)\mathbf{e}_L(l).$$

Let us first remark that this new formulation allows to reduce the usual cost of computation of a DVM exactly to $O(N^{2d})$ (as with the usual spectral method). Note however that the $(2N+1)^d \times (2N+1)^d$ matrix of coefficients $(\beta(K,L))_{K,L}$ has to be computed and stored first, thus the storage requirements are larger with respect to usual DVM.

Now the aim is to give an expansion of $\beta(K,L)$ of the form

$$\beta_{K,L} \simeq \sum_{p=1}^{M} \alpha_p(K)\, \alpha'_p(L)$$

to get a lower cost by the use of discrete convolution.

2.5.2 Expansion of the discrete kernel modes

We make a decoupling assumption as in the spectral case

$$\tilde{B}(|k|,|l|)\, w_{k,l} = a(k)\, b(l).$$

Remark 5 *Note that the DVM constructed by quadrature, in dimension* 3 *for hard spheres, in Panferov and Heintz (2002) satisfies this decoupling assumption with* $a(k) = h^4/gcd(k_1,k_2,k_3)$ *and* $b(l) = 1$ *(see (Panferov and Heintz, 2002, Formula (2.8))), and* $gcd(k_1,k_2,k_3)$ *denotes the greater common divisor of the three integers.*

The difference here with the spectral method, which is a continuous numerical method, is that we have to *enumerate* the set of $\{-\tilde{N} \leq k, l \leq \tilde{N} \mid k \perp l\}$. This motivates for a detailed study of the number of lines passing through 0 and another point in the grid. To this purpose let us introduce the Farey series and a new parameter $0 \leq \bar{N} \leq \tilde{N}$ for the size of the grid used to compute the number of directions. The usual Farey serie is

$$\mathcal{F}^1_{\bar{N}} = \big\{(p,q) \in [|0,\bar{N}|]^2 \mid 0 \leq p \leq q \leq \bar{N} \text{ and } \gcd(p,q) = 1\big\}$$

where $\gcd(p,q)$ denotes again the greater common divisor of the two integers (more details can be found in Hardy and Wright (1979)). It is straightforward to see that the number of lines $A^1_{\bar{N}}$ passing through 0 in the grid $[|-\bar{N},\bar{N}|]^2$ is $A^1_{\bar{N}} = 4\,|\mathcal{F}^1_{\bar{N}}|$. In fact symmetries often allow to reduce the number of directions needed. Similarly one can define

$$\mathcal{F}^2_{\bar{N}} = \big\{(p,q,r) \in [|0,\bar{N}|]^3 \mid 0 \leq p \leq q \leq r \leq \bar{N} \text{ and } \gcd(p,q,r) = 1\big\}$$

and the number of lines $A^2_{\bar{N}}$ passing through 0 in the grid $[|-\bar{N},\bar{N}|]^3$ is $A^2_{\bar{N}} = 16\,|\mathcal{F}^2_{\bar{N}}|$. The exponents of the Farey series refer to the dimension of the space of lines (which is $d-1$). Now let us estimate the cardinal of $\mathcal{F}^1_{\bar{N}}$ and $\mathcal{F}^2_{\bar{N}}$ (see Mouhot and Pareschi (2011) for details).

Lemma 3 *The Farey series in dimension* $d=2$ *and* $d=3$ *satisfy the following asymptotic behavior*

$$\begin{aligned} |\mathcal{F}^1_{\bar{N}}| &= \frac{\bar{N}^2}{2\,\zeta(2)} + O(\bar{N}\,\log\bar{N}) = \frac{3\bar{N}^2}{\pi^2} + O(\bar{N}\,\log\bar{N}) \\ |\mathcal{F}^2_{\bar{N}}| &= \frac{\bar{N}^3}{4\,\zeta(3)} + O(\bar{N}^2) \end{aligned}$$

where ζ *denotes the usual zeta function.*

Now one can deduce the following decomposition of the kernel modes

$$\begin{aligned}\beta(K,L) &= \sum_{-\tilde{N}\leq k,l\leq \tilde{N}} \mathbf{1}_{(k\cdot l)}\, a(|k|)\, b(|l|)\, e_K(k) e_L(l) \\ &\simeq \sum_{e\in\mathcal{A}_{\bar{N}}} \left[\sum_{k\in e\mathbb{Z},\, -\tilde{N}\leq k\leq \tilde{N}} a(|k|)\, e_K(k)\right]\left[\sum_{l\in e^{\perp},\, -\tilde{N}\leq l\leq \tilde{N}} b(|l|)\, e_L(l)\right]\end{aligned}$$

with equality if $\bar{N}=\tilde{N}$. Here $\mathcal{A}_{\bar{N}}$ denotes the set of primal representants of directions of lines in $[|-\bar{N},\bar{N}|]$ passing through 0. After indexing this set, which has cardinal $A^d_{\bar{N}}$, one gets

$$\beta_{K,L} \simeq \sum_{p=1}^{A^d_{\bar{N}}} \alpha_p(K)\, \alpha'_p(L) \tag{50}$$

with

$$\alpha_p(K) = \sum_{k\in e_p\mathbb{Z},\, -\tilde{N}\leq k\leq \tilde{N}} a(|k|)\, e_K(k), \qquad \alpha'_p(L) = \sum_{l\in e_p^{\perp},\, -\tilde{N}\leq l\leq \tilde{N}} b(|l|)\, e_L(l).$$

2.5.3 Implementation aspects

The method yields a decomposition of the discrete collision operator

$$D^{\tilde{N}} \simeq D^{\tilde{N},\bar{N}} = \sum_{p=1}^{A^d_{\bar{N}}} D^{\tilde{N},\bar{N},p}$$

with equality if $\bar{N}=\tilde{N}$. Each $D^{\tilde{N},\bar{N},p}(f,f)$ is defined by the p-th term of the decomposition of the kernel modes (50). Each term $D^{\tilde{N},\bar{N},p}$ of the sum is a discrete convolution operator when it is written in Fourier space.

Thus one can see that even if we take $\bar{N}=\tilde{N}=N$, i.e we take all possible directions in the grid $[|-N,N|]^d$, we get the computational cost $O(N^{2d}\log_2 N)$ which is better than the usual cost of the DVM, $O(N^{2d+1})$ (but slightly worse than the cost $O(N^{2d})$ obtained by solving directly the pseudo-spectral scheme, thanks to a bigger storage requirement).

More generally for a choice of $\bar{N}<N$ we obtain the cost $O(\bar{N}^d N^d \log_2 N)$. The same remarks we did for the fast spectral algorithms about the parallelization and adaptivity (and storage interest) of the method hold true in this case: a parallel algorithm could reduce the computational cost up to $O(N^d \log_2 N)$.

Moreover we expect that for DVM one can strongly reduce the parameter $\bar{N}$ in order to improve the cost of the scheme, without damaging the accuracy of the scheme. The justification for this is the low accuracy of the method (the reduction of the number of direction has a small effect on the already poor accuracy of the scheme).

2.6 Numerical results

In this section we will present several numerical results for the space homogeneous equation which show the improvement of the fast spectral algorithms with respect to the classical spectral methods. The time discretization is performed by standard explicit Runge-Kutta methods.

2.6.1 2D Maxwell molecules

Comparison to exact solutions

We consider 2D pseudo-Maxwell molecules (*i.e.*, the VHS model with $\gamma = 0$). In this case we have an exact solution given by

$$f(t, v) = \frac{\exp(-v^2/2S)}{2\pi\, S^2} \left[2\, S - 1 + \frac{1 - S}{2\, S}\, v^2 \right]$$

with $S = 1 - \exp(-t/8)/2$, which corresponds to the well known "BKW" solution (Bobylev, 1975). This test is performed to check spectral accuracy, by comparing the error at a given time, when using $n_v = 8$, 16 and 32 Fourier modes for each coordinate. We present the results obtained by the classical spectral method and the fast spectral method with different numbers of discrete angles.

Figure 7 shows the relative L^∞, L^1, and L^2 norms of the difference between the numerical and the exact solution, as a function of time. We refer to Filbet et al. (2006) for a more detailed discussion about the different source of error.

Concerning the comparison between the classical and fast spectral methods, we observe that for a fixed value of n_v, the numerical error of the classical spectral method and of the fast algorithm is of the same order. Moreover, the influence of the number of discrete angles is very weak. In Table 1, we give a quantitative comparison of the numerical error $\mathcal{E}_1$ at time $T_{end} = 1$. We can also observe the spectral accuracy for the classical and fast methods: the order of accuracy is about 3 between 8 and 16 grid points, whereas it becomes 7 between 16 to 32 points.

Efficiency and accuracy

Now, we still consider 2D pseudo-Maxwell molecules (*i.e.*, $\gamma = 0$) with the following initial datum

$$f(0, v) = \frac{1}{4\,\pi} \left[\exp\left(-\frac{|v - v_0|^2}{2} \right) + \exp\left(-\frac{|v + v_0|^2}{2} \right) \right], \quad v \in \mathbb{R}^2,$$

where $v_0 = (1, 2)$. In this case, we do not know the exact solution but we want to study the influence of the number of discrete angles on a non-isotropic solution. Thus, this test is used to check the energy conservation and the evolution of high-order moments of the solution. Simulations are performed with n_v=16, 32 and 64 points.

Number of points	Classical spectral	Fast spectral with $M = 4$	Fast spectral with $M = 6$	Fast spectral with $M = 8$
8	0.02013	0.02778	0.02129	0.02112
16	0.00204	0.00329	0.00238	0.00224
32	1.405E-5	2.228E-5	1.861E-5	1.772E-5

Table 1: Comparison of the L^1 error in $2D$ between the classical spectral method and the fast spectral method with different numbers of discrete angles and with a second-order Runge-Kutta time discretization at time $T_{end} = 1$.

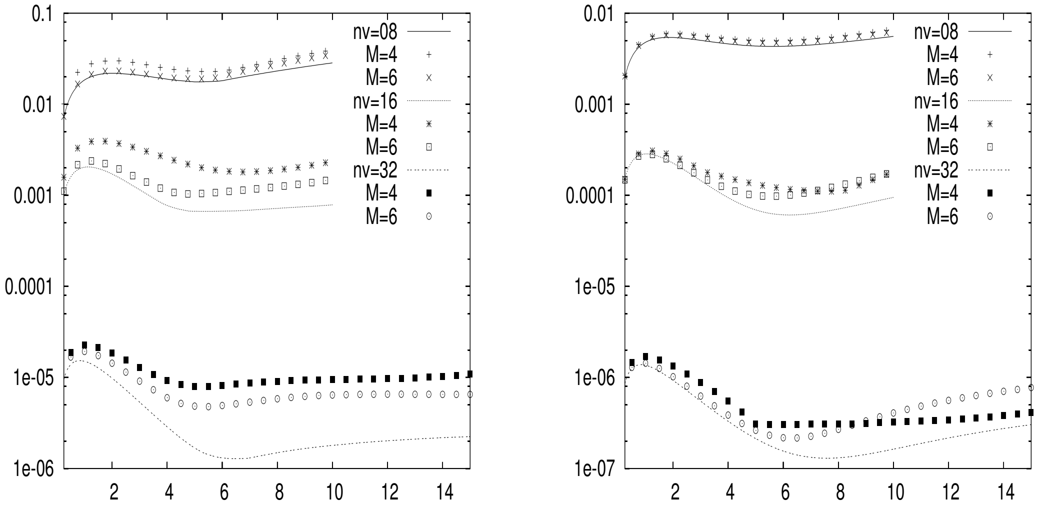


Figure 7: Evolution of the numerical L^1 and L^∞ relative error of $f(t,v)$.

In Figure 8 the relaxation of the entropy and the temperature components for the fast and classical spectral methods is shown. Finally, we plot in Figure 9 the time evolution of high-order moments of $f_N(t,v)$ given in discrete form by

$$\mathcal{M}_k(t) = \Delta v^2 \sum_{l=-N}^{N} |v_l|^k f_N(t,v_l).$$

High-order moments give information on the accuracy of the approximate distribution function tail. Once again, we observe that the number of angles does not affect the results even if the solution is non-isotropic.

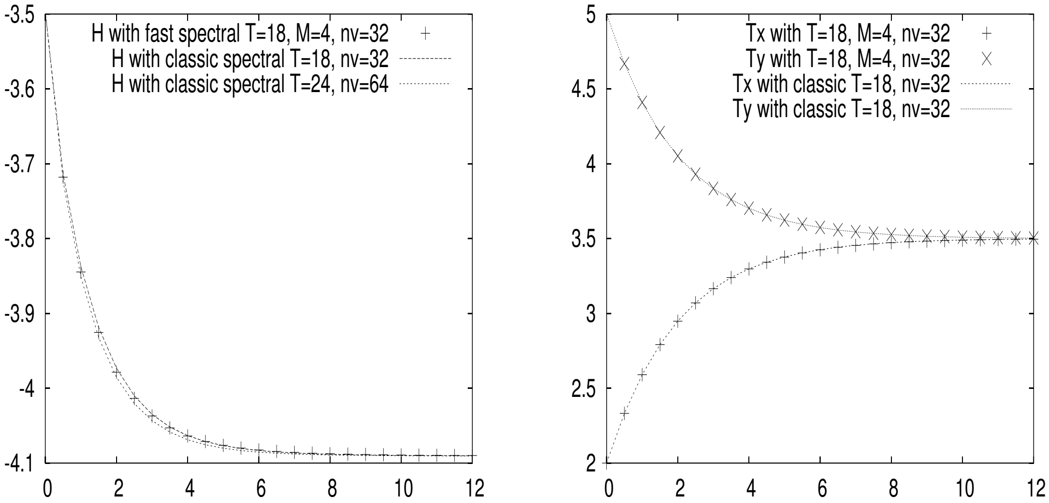


Figure 8: Relaxation of the entropy and the temperature components for the fast and classical spectral methods with respect to the number of modes per direction n_v and the length box T.

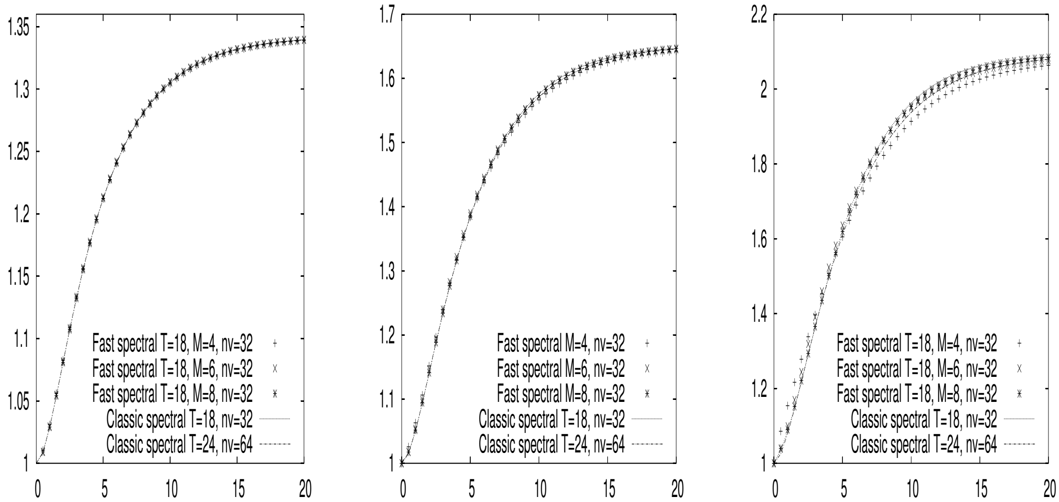


Figure 9: Time evolution of the variations of high order normalized moments $\mathcal{M}_4$, $\mathcal{M}_5$ and $\mathcal{M}_6$ of $f(t, v)$ for the fast and classical spectral methods with respect to the number of modes per direction n_v and the length box T.

2.6.2 3D Hard Spheres

In this section we consider the 3D Hard Sphere molecules (HS) model. The initial condition is chosen as the sum of two Gaussians

$$f(v, 0) = \frac{1}{2(2\pi\sigma^2)} \left[\exp\left(-\frac{|v - v_0|^2}{2\sigma^2} \right) + \exp\left(-\frac{|v + v_0|^2}{2\sigma^2} \right) \right]$$

with $\sigma = 1$ and $v_0 = (2, 1, 0)$. The final time of the simulation is $T_{end} = 3$ and corresponds approximatively to the time for which the steady state of the solution is reached. The time step is $\Delta t = 0.1$ and the length box is taken as $T = 12$ when $n_v = 16$ and $T = 15$ when $n_v = 32$.

This test is used to check the evolution of moments and particularly the stress tensor $P_{i,j}$, $i, j = 1, \cdots, 3$ defined as

$$P_{i,j} = \int_{\mathbb{R}^3} f(v)(v_i - u_i)\,(v_j - u_j)\, dv, \quad (i, j) \in \{1, 2, 3\}^2,$$

where $(u_i)_i$ are the components of the mean velocity. In Figure 10, we propose the evolution of the temperature for the two methods using 32 grid points in each direction. The solution is also compared with the solution obtained from a standard Direct Simulation Monte-Carlo method. We remark that in dimension $d = 3$ the speed-up of the methods becomes really evident for large values of N. For example, for $N = 64$ and $M = 4$ the fast method is more than 14 times faster then the direct algorithm.

3 Asymptotic preserving splitting methods

In this section we will consider the problem of numerically solving the full nonhomogeneous equation. The approximation of the velocity space has been extensively discussed

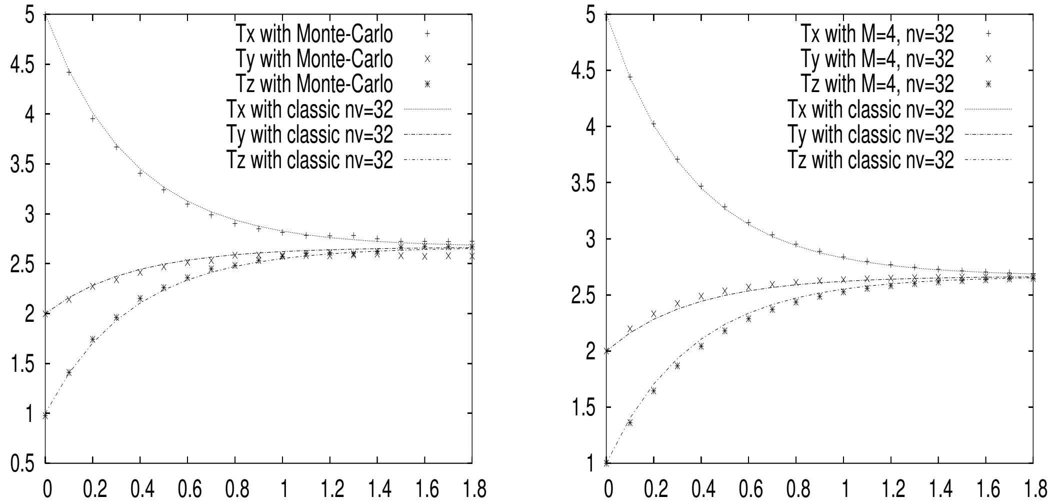


Figure 10: Comparison between the fast and classical spectral methods and the Monte-Carlo methods for the temperature components relaxation.

in Section 2, thus here we will focus on the space and time discretizations. Besides the solution of the transport part, when dealing with the full problem the main difficulty is the construction of schemes which are robust in the fluid limit, or, as defined in Section 1.7, which are asymptotic preserving (AP). Of course AP is an important property strictly related to the stability of the time discretization scheme in stiff regimes. Here we will treat separately methods based on operator splitting from other time discretizations which avoid operator splitting. First we start from operator splitting methods which have been introduced in Section 1.6. In the next paragraphs we will focus on the space-time approximation of the transport step and on the time discretization of the collision step.

3.1 Transport step

The solution of the transport equation can be obtained in a variety of ways accordingly to the particular application considered (see for example LeVeque (1992); Filbet et al. (2001); Roe and Sidilkover (1992) and the references therein). A review of such a broad field is above the scope of this paper. Here we give the details of the methods developed originally in Filbet et al. (2001) which has several nice properties, among which to allow the use of large time steps even for large velocities.

3.1.1 Positive and Conservative schemes

Let us consider the transport equation written as

$$\partial_t f + \nabla_x (v\, f) = 0, \quad \forall (t,x) \in I\!R^+ \times I\!R^d. \tag{51}$$

Then, the solution of the transport equation at time t^{n+1} reads

$$f(t^{n+1}, x) = f(t^n, x - v\,\Delta t), \quad \forall x \in I\!R^d.$$

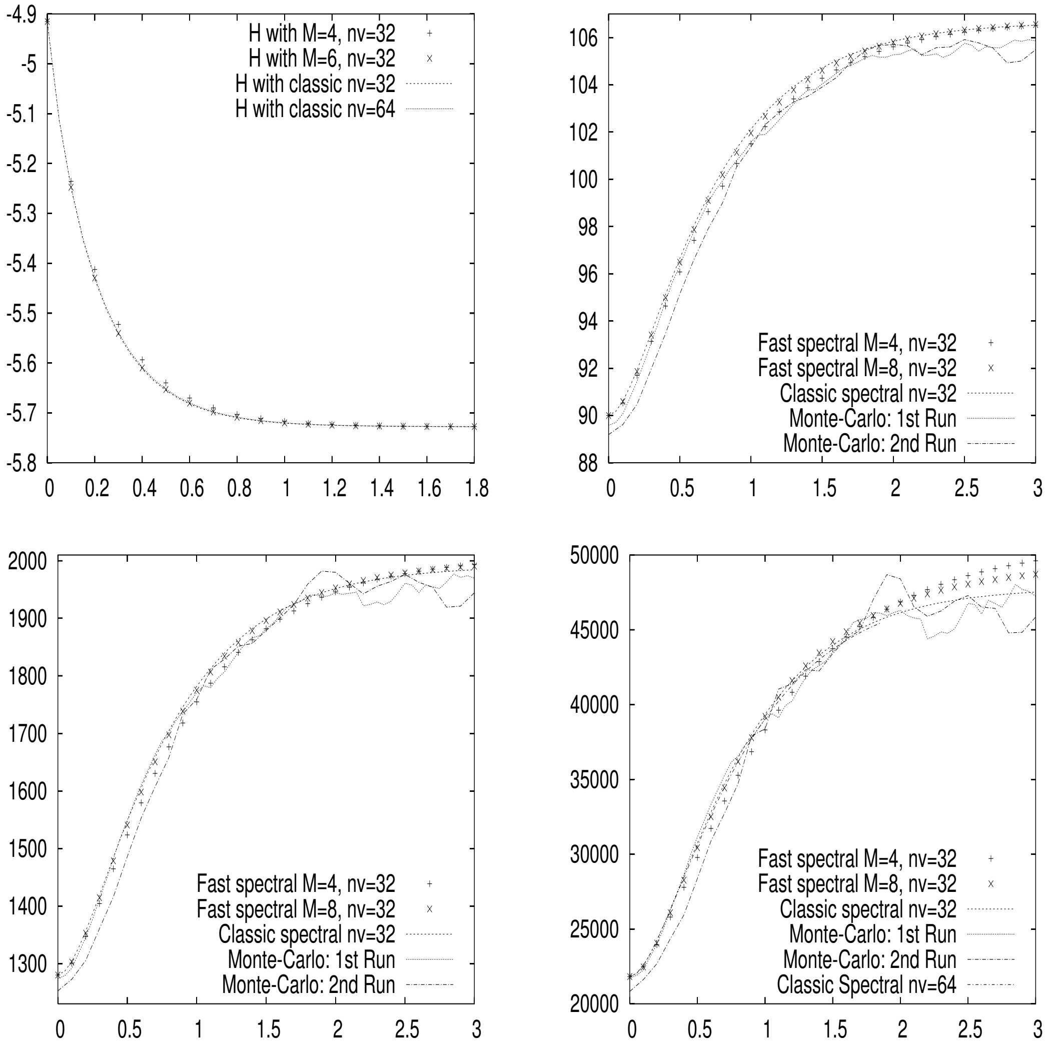


Figure 11: Time evolution of the kinetic entropy H and high-order moments $\mathcal{M}_4$, $\mathcal{M}_6$ and $\mathcal{M}_8$ of $f(t, v)$ for the fast and classical spectral methods, and the Monte-Carlo methods.

For simplicity, let us restrict ourselves to a one dimensional problem and introduce a finite set of mesh points $\{x_{i+1/2}\}_{i\in I}$ on the computational domain. We will use the notations $\Delta x = x_{i+1/2} - x_{i-1/2}$, $C_i = [x_{i-1/2}, x_{i+1/2}]$ and x_i the center of C_i. Assume the values of the distribution function are known at time $t^n = n\,\Delta t$ on cells C_i, we compute the new values at time t^{n+1} by integration of the distribution function on each sub-interval. Thus, using the explicit expression of the solution, we have

$$\int_{x_{i-1/2}}^{x_{i+1/2}} f(t^{n+1}, x)dx = \int_{x_{i-1/2}-v\,\Delta t}^{x_{i+1/2}-v\,\Delta t} f(t^n, x)dx,$$

then, setting

$$\Phi_{i+1/2}(t^n) = \int_{x_{i+1/2}-v\,\Delta t}^{x_{i+1/2}} f(t^n, x)dx,$$

we obtain the conservative form

$$\int_{x_{i-1/2}}^{x_{i+1/2}} f(t^{n+1}, x)dx = \int_{x_{i-1/2}}^{x_{i+1/2}} f(t^n, x)dx \,+\, \Phi_{i-1/2}(t^n) \,-\, \Phi_{i+1/2}(t^n). \tag{52}$$

The evaluation of the average of the solution over $[x_{i-1/2}, x_{i+1/2}]$ allows to ignore fine details of the exact solution which may be costly to compute. The main step is now to choose an efficient method to reconstruct the distribution function from the cell average on each cell C_i. We will consider a reconstruction via primitive function preserving positivity and maximum values of f (Filbet et al., 2001). Let $F(t^n, x)$ be a primitive of the distribution function $f(t^n, x)$, if we denote by

$$f_i^n = \frac{1}{\Delta x}\int_{x_{i-1/2}}^{x_{i+1/2}} f(t^n, x)dx,$$

then $F(t^n, x_{i+1/2}) - F(t^n, x_{i-1/2}) = \Delta x\, f_i^n$ and

$$F(t^n, x_{i+1/2}) = \Delta x \sum_{k=0}^{i} f_k^n = w_i^n.$$

First we construct an approximation of the primitive on the small interval $[x_{i-1/2}, x_{i+1/2}]$ using the stencil $\{x_{i-3/2}, x_{i-1/2}, x_{i+1/2}, x_{i+3/2}\}$

$$\begin{aligned}\tilde{F}_h(t^n, x) \;&=\; w_{i-1}^n + (x - x_{i-1/2})f_i^n + \frac{1}{2\Delta x}(x - x_{i-1/2})(x - x_{i+1/2})[f_{i+1}^n - f_i^n]\\ &+\; \frac{1}{6\Delta x^2}(x - x_{i-1/2})(x - x_{i+1/2})(x - x_{i+3/2})[f_{i+1}^n - 2\,f_i^n + f_{i-1}^n],\end{aligned}$$

where we use the relation $w_i^n - w_{i-1}^n = \Delta x\, f_i^n$. Thus, by differentiating $\tilde{F}_h(x)$, we obtain a third order accurate approximation of the distribution function on the interval $[x_{i-1/2}, x_{i+1/2}]$

$$\begin{aligned}\tilde{f}_h(t^n, x) = \frac{\partial \tilde{F}_h}{\partial x}(t^n, x) = f_i^n +&\\ +\; \frac{1}{6\,\Delta x^2}&\Big[2\,(x - x_i)(x - x_{i-3/2}) + (x - x_{i-1/2})(x - x_{i+1/2})\Big](f_{i+1}^n - f_i^n)\\ -\; \frac{1}{6\,\Delta x^2}&\Big[2\,(x - x_i)(x - x_{i+3/2}) + (x - x_{i-1/2})(x - x_{i+1/2})\Big](f_i^n - f_{i-1}^n).\end{aligned}$$

Unfortunately, this approximation does not preserve positivity of the distribution function f. Then, in order to satisfy a maximum principle and to avoid spurious oscillations we introduce slope correctors

$$\begin{aligned} f_h(t^n,x) &= f_i^n + \\ &+ \frac{\varepsilon_i^+}{6\,\Delta x^2}\Big[2\,(x-x_i)(x-x_{i-3/2}) + (x-x_{i-1/2})(x-x_{i+1/2})\Big](f_{i+1}^n - f_i^n) \\ &- \frac{\varepsilon_i^-}{6\,\Delta x^2}\Big[2\,(x-x_i)(x-x_{i+3/2}) + (x-x_{i-1/2})(x-x_{i+1/2})\Big](f_i^n - f_{i-1}^n), \end{aligned} \tag{53}$$

with

$$\varepsilon_i^{\pm} = \begin{cases} \min\Big(1, 2\,f_i^n/(f_{i\pm1}^n - f_i^n)\Big) & \text{if } f_{i\pm1}^n - f_i^n > 0, \\ \\ \min\Big(1, -2\,(f_\infty - f_i^n)/(f_{i\pm1}^n - f_i^n)\Big) & \text{if } f_{i\pm1}^n - f_i^n < 0, \end{cases} \tag{54}$$

where $f_\infty = \max\limits_{j\in I}\{f_j^n\}$ is a local maximum.

The theoretical properties of this reconstruction can be summarized by the following(Filbet et al., 2001)

Proposition 2 *The approximation of the distribution function $f_h(x)$, defined by (53)-(54), satisfies*

- *The conservation of the average: for all $i \in I$, $\int_{x_{i-1/2}}^{x_{i+1/2}} f_h(x)dx = \Delta x\, f_i$.*
- *The maximum principle: for all $x \in (x_{min}, x_{max})$, $0 \le f_h(x) \le f_\infty$.*

Moreover, if we assume that the Total Variation of the distribution function $f(x)$ is bounded, then we obtain the global estimate

$$\int_{x_{min}}^{x_{max}} |f_h(x) - \tilde{f}_h(x)|\,dx \le 4\,TV(f)\,\Delta x,$$

where $\tilde{f}_h$ denotes the third order approximation of f without slope corrector.

Remark 6 *If the solution is smooth, we can check numerically that the scheme is third order. In several dimensions we can perform reconstruction dimension by dimension using tensor product.*

3.2 Time discretization of the collision step

Since we aim at developing operator splitting AP schemes, the most natural choice would be to use implicit solvers applied to the collision step (32). Unfortunately the use of fully implicit schemes for the full Boltzmann collision integral is unpracticable due to the prohibitive computational cost required by the solution of the very large non-linear algebraic system originated by the five fold integral appearing in $Q(f,f)$. Exponential methods represent a possible way to overcome these difficulties.

3.2.1 Problem reformulation

First we rewrite the homogeneous equation (32) in the form

$$\partial_t f = \frac{1}{\varepsilon}(P(f,f) - \mu f), \tag{55}$$

where $P(f,f) = Q(f,f) + \mu f$ and $\mu > 0$ is a constant such that $P(f,f) \geq 0$. Typically μ is an estimate of the largest spectrum of the loss part of the collision operator.

By construction we have the following

$$\frac{1}{\mu}\int_{\mathbb{R}^3} P(f,f)(v)\phi(v)\,dv = \int_{\mathbb{R}^3} f(v)\phi(v)\,dv, \quad \phi(v) = 1, v, v^2. \tag{56}$$

Thus $P(f,f)/\mu$ is a density function and we can consider the following decomposition

$$P(f,f)/\mu = M + g, \tag{57}$$

where M is the Maxwellian with the same moments of f.

The function g represents the non equilibrium part of the distribution function and from the definition above it follows that g is in general non positive. Moreover since $P(f,f)/\mu$ and M have the same moments we have

$$\int_{\mathbb{R}^3} g(v)\phi(v)\,dv = 0, \quad \phi(v) = 1, v, v^2. \tag{58}$$

The homogeneous equation can be written in the form

$$\partial_t f = \frac{\mu}{\varepsilon} g + \frac{\mu}{\varepsilon}(M - f) = \frac{\mu}{\varepsilon}\left(\frac{P(f,f)}{\mu} - M\right) + \frac{\mu}{\varepsilon}(M - f). \tag{59}$$

The above system is equivalent to the penalization method for the collision operator recently introduced in Filbet and Jin (2010). Note that even if M is nonlinear in f, thanks to the conservation properties (19), it remains constant during the relaxation process. The main feature of such formulation is that on the right hand side we have a stiff dissipative linear part $\mu(M - f)/\varepsilon$ which characterizes the asymptotic behavior of f and a stiff non dissipative non linear part $(P(f,f)/\mu - M)/\varepsilon$ which describes the deviations of $P(f,f)/\mu$ from M, or equivalently the deviations of the Boltzmann operator from a BGK-like relaxation term.

The formulation above permits to use exponential integrators where the schemes take advantage of the exact solution of the linear part (Dimarco and Pareschi, 2010a). Exponential methods for kinetic equations were first proposed in Gabetta et al. (1997).

3.2.2 Explicit exponential Runge-Kutta schemes

In order to derive the methods it is useful to rewrite (59) as

$$\frac{\partial (f - M)e^{\mu t/\varepsilon}}{\partial t} = \frac{1}{\varepsilon}(P(f,f) - \mu M)e^{\mu t/\varepsilon}. \tag{60}$$

The above form is readily obtained if one multiplies (59) by the integrating factor $\exp(\mu t/\varepsilon)$ and takes into account the fact that M does not depend of time. A class of explicit exponential Runge-Kutta schemes is then obtained by direct application of an explicit Runge-Kutta method to (60). More in general we can consider the family of methods characterized by

$$\begin{aligned} F^{(i)} &= e^{-c_i \mu\Delta t/\varepsilon} f^n + \frac{\mu\Delta t}{\varepsilon} \sum_{j=1}^{i-1} A_{ij}(\mu\Delta t/\varepsilon) \left(\frac{P(F^{(j)}, F^{(j)})}{\mu} - M^n \right) \\ &+ \left(1 - e^{-c_i \mu\Delta t/\varepsilon}\right) M^n, \qquad i = 1, \dots, \nu \end{aligned} \tag{61}$$

$$\begin{aligned} f^{n+1} &= e^{-\mu\Delta t/\varepsilon} f^n + \frac{\mu\Delta t}{\varepsilon} \sum_{i=1}^{\nu} W_i(\mu\Delta t/\varepsilon) \left(\frac{P(F^{(i)}, F^{(i)})}{\mu} - M^n \right) \\ &+ \left(1 - e^{-\mu\Delta t/\varepsilon}\right) M^n, \end{aligned} \tag{62}$$

where Δt is the time step, $f^n = f(t^n)$, $M^n = M(t^n)$, $c_i \geq 0$, and the coefficients A_{ij} and the weights W_i are such that

$$A_{ij}(0) = a_{ij}, \quad W_i(0) = w_i, \quad i, j = 1, \dots, \nu$$

with coefficients a_{ij} and weights w_i given by a standard explicit Runge-Kutta method called the underlying method. Various schemes come from the different choices of the underlying method. The most popular approaches are the integrating factor (IF) and the exponential time differencing (ETD) methods (Maset and Zennaro, 2009). Since M^n does not depend on time during the collision process in the sequel we will omit the index n.

For the so-called Integrating Factor methods, which correspond to a direct application of the underlying method to (60), we have

$$\begin{aligned} A_{ij}(\lambda) &= a_{ij} e^{-(c_i - c_j)\lambda}, \quad i, j = 1, \dots, \nu, \quad i > j \\ W_i(\lambda) &= w_i e^{-(1-c_i)\lambda}, \quad i = 1, \dots, \nu, \end{aligned} \tag{63}$$

with $\lambda = \mu\Delta t/\varepsilon$.

The first order IF scheme reads

$$f^{n+1} = e^{-\frac{\mu\Delta t}{\varepsilon}} f^n + \frac{\mu\Delta t}{\varepsilon} e^{-\frac{\mu\Delta t}{\varepsilon}} \left(\frac{P(f^n, f^n)}{\mu} - M \right) + \left(1 - e^{-\frac{\mu\Delta t}{\varepsilon}}\right) M, \tag{64}$$

which is based on explicit Euler. For such methods the order of accuracy is the same as the order of the underlying method.

The Exponential Time Differencing methods are strictly connected with the integral representation of (60). In the general case the coefficients for ETD methods have the form

$$\begin{aligned} A_{ij}(\lambda) &= \int_0^1 e^{(1-s)c_i\lambda} p_{ij}(s)\, ds, \quad i, j = 1, \dots, \nu, \quad i > j \\ W_i(\lambda) &= \int_0^1 e^{(1-s)\lambda} p_i(s)\, ds, \quad i = 1, \dots, \nu, \end{aligned}$$

where p_i and p_{ij} are suitable polynomials.

The standard first order ETD method based on explicit Euler in our case gives

$$f^{n+1} = e^{-\frac{\mu\Delta t}{\varepsilon}} f^n + \frac{\mu\Delta t}{\varepsilon} \varphi\left(\frac{\mu\Delta t}{\varepsilon}\right) \frac{P(f^n, f^n)}{\mu}, \tag{65}$$

where $\varphi(z) = (1 - e^{-z})/z$.

3.2.3 Time Relaxed methods

A class of exponential methods for kinetic equations, the so-called time relaxed (TR) methods, has been introduced in Gabetta et al. (1997) as a combination of an exponential expansion (or Wild sum) together with a suitable Maxwellian truncation. In this paragraph we show that these schemes included already decomposition (59) and can be derived directly from a suitable Taylor expansion of (60).

To show this, let us first introduce the change of variables

$$\tau = 1 - \exp(-\mu t/\varepsilon),$$

which, using the bilinearity of $P(f,f)$, gives the equation

$$\frac{\partial}{\partial\tau}\left[(f-M)\frac{1}{1-\tau}\right] = (P(f,f)-\mu M)\frac{1}{\mu(1-\tau)^2}. \tag{66}$$

The application of an explicit Runge-Kutta scheme to (66) with time step $\Delta\tau = 1 - \exp(-\mu\Delta t/\varepsilon)$ leads to a class of ETD methods. For example the first order scheme based on explicit Euler in the original variables yields

$$f^{n+1} = e^{-\frac{\mu\Delta t}{\varepsilon}} f^n + \frac{\mu\Delta t}{\varepsilon}\varphi_1\left(\frac{\mu\Delta t}{\varepsilon}\right)\left(\frac{P(f^n,f^n)}{\mu} - M\right) + (1-e^{-\frac{\mu\Delta t}{\varepsilon}})M, \tag{67}$$

where $\varphi_k(z) = e^{-z}(1-e^{-z})^k/z$, $k = 1,2,\ldots$.

Note that such scheme coincides with the first order exponential time relaxed method derived in Gabetta et al. (1997) and differs from the standard ETD method based on explicit Euler. Higher order ETD methods can be derived as well simply applying higher order explicit Runge-Kutta methods to (66). Although interesting, here we do not explore further this class of schemes.

Now let us consider a different approach by taking the Taylor expansion of $(f-M)/(1-\tau)$ around $\tau = 0$ in (66). This leads to

$$\begin{aligned}(f-M)/(1-\tau) &= (f_0 - M) + \tau\left[\frac{P(f_0,f_0)}{\mu} - M\right] \\ &+ \frac{\tau^2}{2}\left[\frac{P(P(f_0,f_0),f_0) + P(f_0,P(f_0,f_0))}{\mu^2} - 2M\right] + O(\tau^3)\end{aligned}$$

where we have used the bilinearity of the operator $P(f,f)$.

If we compute all the terms in the expansion and use recursively the bilinearity of $P(f,f)$ we can state the following

Proposition 3 *The solution to problem (59) or equivalently (60) or (66) can be represented as*

$$f(v,t) = (1-\tau)f_0(v) + (1-\tau)\sum_{k=1}^{\infty}\tau^k(f_k^n(v) - M(v)) + \tau M(v), \tag{68}$$

where the functions f_k are given by the recurrence formula

$$f_{k+1}(v) = \frac{1}{k+1}\sum_{h=0}^{k}\frac{1}{\mu}P(f_h, f_{k-h})(v), \quad k = 0,1,\ldots. \tag{69}$$

By truncating expansion (68) at the order m, and reverting to the old variables, we recover exactly the time relaxed schemes presented in Gabetta et al. (1997)

$$f^{n+1} = e^{-\mu\Delta t/\varepsilon} f^n + e^{-\mu\Delta t/\varepsilon} \sum_{k=1}^{m} (1 - e^{-\mu\Delta t/\varepsilon})^k (f_k^n - M) + (1 - e^{-\mu\Delta t/\varepsilon})M, \tag{70}$$

which, using the fact that

$$1 - e^{-\mu\Delta t/\varepsilon} \sum_{k=0}^{m} (1 - e^{-\mu\Delta t/\varepsilon})^k = (1 - e^{-\mu\Delta t/\varepsilon})^{m+1},$$

can be rewritten in the usual form emphasizing their convexity properties

$$f^{n+1} = e^{-\mu\Delta t/\varepsilon} \sum_{k=0}^{m} (1 - e^{-\mu\Delta t/\varepsilon})^k f_k^n + (1 - e^{-\mu\Delta t/\varepsilon})^{m+1} M. \tag{71}$$

A remarkable feature of these methods is that the functions $f_k(v)$ are density functions with the same moments of the initial data. Such property, together with unconditional nonnegativity and convexity of the weights, is enough to guarantee asymptotic preservation of the schemes as well as nonnegativity and entropic stability (see Gabetta et al. (1997) for details).

3.2.4 Main properties

In this section we will report the main properties for an IF exponential scheme in the form (61)-(62). We refer to Dimarco and Pareschi (2010a) for further details an results concerning general exponential schemes.

Now let us denote by f^n and g^n the corresponding solutions obtained with an explicit exponential Runge-Kutta method. Let us define

$$R(\lambda) \;\; = \;\; e^{-\lambda} + \sum_{k=0}^{\nu-1} \lambda^{k+1} \bar{w}(\lambda)^T \bar{A}(\lambda)^k \bar{E}(\lambda)\bar{e}, \tag{72}$$

where $\lambda = \mu\Delta t/\varepsilon$, $\bar{A}(\lambda)$ the $\nu \times \nu$ matrix of elements $|A_{ij}(\lambda)|$, $\bar{w}(\lambda)$ the $\nu \times 1$ vector of elements $|W_i(\lambda)|$, $\bar{e}$ the $\nu \times 1$ unit vector and $\bar{E}(\lambda) = \text{diag}(e^{-c_1\lambda}, \ldots, e^{-c_\nu\lambda})$.

We can state (Dimarco and Pareschi, 2010a)

Theorem 3 *If an explicit exponential Runge-Kutta method in the form (61)-(62) satisfies*

$$\lim_{\lambda\to\infty} R(\lambda) = 0, \tag{73}$$

with $R(\lambda)$ given by (72) then it is asymptotic preserving.

Note that for an IF method we have

$$|A_{ij}(\lambda)| \leq |a_{ij}| e^{-(c_i - c_j)\lambda}, \quad |W_i(\lambda)| \leq |w_i| e^{-(1-c_i)\lambda},$$

thus we require

$$0 = c_1 \leq c_2 \ldots \leq c_\nu \leq 1, \tag{74}$$

in order for the above quantities to be bounded independently of λ.

Moreover for nonnegative coefficients and weights we get

$$R(\lambda) = e^{-\lambda}\left(1 + \sum_{k=0}^{\nu-1} \lambda^{k+1} w^T A^k \bar{e}\right), \tag{75}$$

and condition (73) is always satisfied.

It can be proved that if the underlying Runge-Kutta method is a ν-stage explicit Runge-Kutta method of order ν, with nonnegative coefficients and weights satisfying (74), then the scheme is unconditionally stable and contractive. As pointed out in Maset and Zennaro (2009) examples of such methods are well-known up to $\nu = 4$ and the classical RK method of order four is the sole method with four stages.

For practical applications it may be convenient to require that as $\lambda \to \infty$ the numerical solution f^{n+1} and each level $F^{(i)}$ of the IF method are projected towards the local Maxwellian. It is straightforward to verify that this stronger AP property is satisfied if we replace condition (74) by

$$0 = c_1 < c_2 < \ldots < c_\nu < 1. \tag{76}$$

We conclude this section with a results concerning an important convexity property of IF schemes. We can state (see Dimarco and Pareschi (2010a))

Proposition 4 *An explicit IF method is unconditionally positive and entropic if the underlying Runge-Kutta method has nonnegative coefficients and weights satisfying By Taylor expansion we obtain conditions*

$$\begin{aligned} \sum_{j=1}^{i-1} a_{ij} {c_j}^k &\leq \frac{c_i^k}{k+1}, \quad k = 0,1,2,\ldots, \quad i = 1,\ldots,\nu \qquad (77)\\ \sum_{i=1}^{\nu} w_i c_i^k &\leq \frac{1}{k+1}, \quad k = 0,1,2,\ldots, \qquad (78) \end{aligned}$$

Note that the above conditions on the choice of the underlying method are quite restrictive since we are not using the bilinearity of $P(f,f)$ which would lead to weaker constraints on a_{ij} and w_i. Examples of methods that satisfy convexity are the second order modified Euler method

$$\begin{aligned} F^{(1)} &= f^n, \\ F^{(2)} &= e^{-\lambda/2} f^n + \frac{\lambda}{2} e^{-\lambda/2}\left(\frac{P(F^{(1)},F^{(1)})}{\mu} - M\right) + \left(1 - e^{-\lambda/2}\right) M, \qquad (79) \\ f^{n+1} &= e^{-\lambda} f^n + \lambda e^{-\lambda/2}\left(\frac{P(F^{(2)},F^{(2)})}{\mu} - M\right) + \left(1 - e^{-\lambda}\right) M. \end{aligned}$$

and the third order Heun method

$$\begin{aligned}
F^{(1)} &= f^n, \\
F^{(2)} &= e^{-\lambda/3} f^n + \frac{\lambda}{3} e^{-\lambda/3} \left(\frac{P(F^{(1)}, F^{(1)})}{\mu} - M \right) + \left(1 - e^{-\lambda/3}\right) M, \\
F^{(3)} &= e^{-2\lambda/3} f^n + \frac{2\lambda}{3} e^{-\lambda/3} \left(\frac{P(F^{(2)}, F^{(2)})}{\mu} - M \right) + \left(1 - e^{-2\lambda/3}\right) M, \qquad (80) \\
f^{n+1} &= e^{-\lambda} f^n + \frac{\lambda}{4} e^{-\lambda} \left(\frac{P(F^{(1)}, F^{(1)})}{\mu} - M \right) \\
&+ \frac{3\lambda}{4} e^{-\lambda/3} \left(\frac{P(F^{(3)}, F^{(3)})}{\mu} - M \right) + \left(1 - e^{-\lambda}\right) M.
\end{aligned}$$

but not the classical fourth order Runge-Kutta scheme.

3.2.5 Implementation aspects

An essential aspect in the reformulation of the problem given by (75) is the choice of the value of the constant μ used in estimating the spectrum of the collision operator. Of course such constant can be chosen at each time step in order to improve our estimate. In the sequel we show different choices in the case of variable hard spheres. We set $C_\gamma = 1$ for simplicity.

The choice of an upper bound for the loss part of the collision term leads to take $\mu = \mu_p$ where

$$\mu_p = \sup_v \int_{\mathbb{R}^3} |v - v_*|^\gamma f(v_*)\, dv_*. \qquad (81)$$

Positivity is guaranteed since it implies clearly $P(f, f) \geq 0$. From a practical viewpoint computation of (81) can be done at $O(N \log N)$ for a deterministic method based on N parameters for representing $f(v)$ on a regular mesh. This can be done using the FFT algorithm thanks to the convolution structure of the loss term in (81).

However, such positivity constraint on $P(f, f)$ typically leads to overestimates of the true spectrum of the collision operator, especially in Monte Carlo simulations. A better estimate of μ would be to use the average collision frequency

$$\mu_a = \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} |v - v_*|^\gamma f(v) f(v_*)\, dv_*\, dv. \qquad (82)$$

Finally, as suggested in Filbet and Jin (2010), μ can be chosen as an estimate of the spectral radius of the linearized operator Q around the Maxwellian M. In fact

$$Q(f, f) \approx Q(M, M) + \nabla Q(M, M)(M - f) = \nabla Q(M)(M - f),$$

where $\nabla Q(M, M)$ is the Frechet derivative of Q evaluated at M. For example one can take

$$\mu_s = \sup_v \left| \frac{Q(f, f)}{f - M} \right|. \qquad (83)$$

The choices $\mu = \mu_a$ or $\mu = \mu_s$, although more accurate, pose the question of stability of the resulting scheme since they do not guarantee $P(f, f) \geq 0$. We refer to Dimarco and Pareschi (2010a) for more results in this direction.

Remark 7 *Here we have assumed μ constant during the time stepping. In principle one can take $\mu = \mu(t)$ and rewrite the exponential methods for a time dependent μ from*

$$\frac{\partial (f - M)e^{\frac{1}{\varepsilon}\int_0^t \mu(s)\,ds}}{\partial t} = \frac{1}{\varepsilon}(P(f,f) - \mu(t)M)e^{\frac{1}{\varepsilon}\int_0^t \mu(s)\,ds}, \tag{84}$$

and then recompute μ_p at each time step or stage of the Runge-Kutta method.

3.3 Numerical tests

In this section we report several test cases. First for homogeneous equations where we illustrate the AP feature of the exponential schemes and then for nonhomogeneous problems in different regimes.

3.3.1 Homogeneous problems

A simple test case

In the first test case we consider the simplified situation of the Kac equation (17) with initial data given by

$$f(v, 0) = v^2 e^{-v^2}.$$

In this case we have an exact solution given by

$$f(v,t) = \frac{1}{2}\left(\frac{3}{2}(1 - C(t))\sqrt{C(t)} + (3C(t) - 1)C(t)^{3/2}v^2\right) e^{-C(t)v^2}, \tag{85}$$

with $C(t) = (3 - 2e^{-\sqrt{\pi}t/16})^{-1}$. Thanks to its simple structure the collision integral can be evaluated analytically and no further approximation is needed when comparing the accuracy of different time discretizations after only one time step. More precisely we compare the first and second order TR and IF methods with a first and second order implicit-explicit (IMEX) and diagonally implicit Runge-Kutta (DIRK) methods. The IMEX methods have been applied following Filbet and Jin (2010) (schemes (4.3) and (4.14)) which in a homogeneous setting correspond to take the gain part of the collision term explicitly and the loss part implicitly since the Maxwellian term cancels. All schemes are unconditionally stable and AP except for the particular IMEX schemes which do not satisfy the AP property. We refer to Section 4.3 for a discussion of asymptotic preserving IMEX schemes.

The results are reported in Figure 12. The AP feature of the exponential schemes (both TR and IF) and fully implicit solvers permits to capture the correct behavior. Quite remarkably, in this test case, exponential schemes are more accurate then the corresponding fully implicit methods with the IF methods being the most accurate.

Accuracy test

Next we compare the accuracy for Maxwell molecules and Hard Spheres in 3D. As initial data we consider an equilibrium distribution with temperature $T = 6$, density $\varrho = 1$ and mean velocity $u = -0.5$. To this distribution we add a bump on the right tail along the x-axis. The bump is realized adding a fraction of particles in equilibrium state with mass

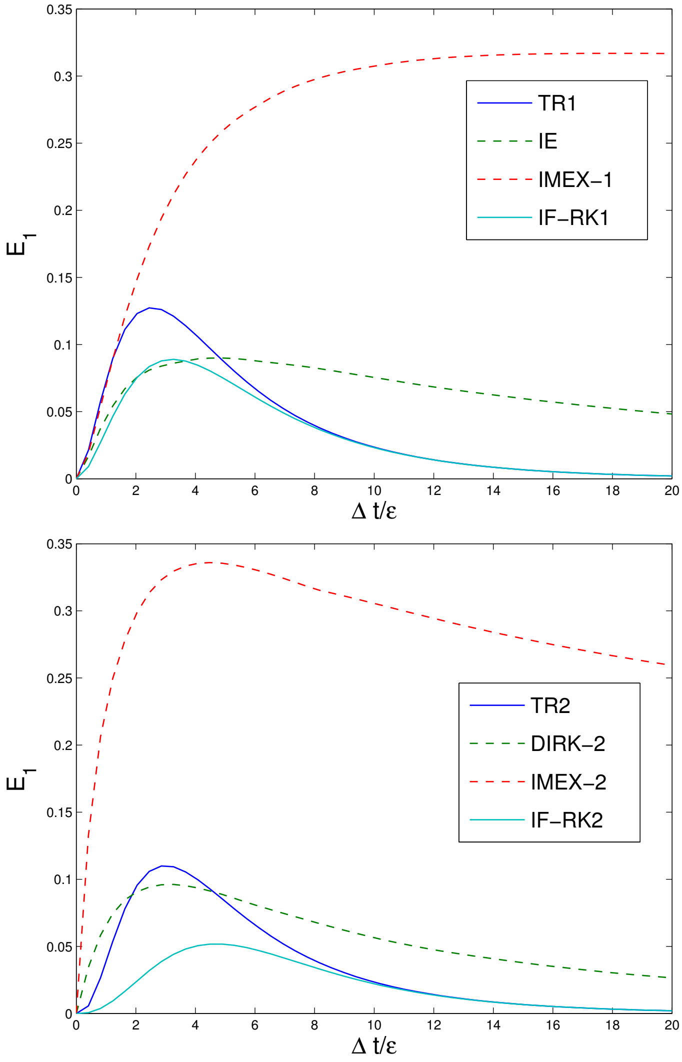


Figure 12: Kac equation L_1 error of some first and second order time discretization methods

$\varrho_b = 0.5\,\varrho$, mean velocity $u_b = 4\sqrt{T}$ and temperature $T_b = 0.5\,T$ to the initial Maxwellian distribution function. The velocity error is neglected by solving the collision operator with Monte Carlo methods with a very large number of particles. The simulations are run till the equilibrium is approximately reached, which means $t = 0.4$ in the case of Hard Spheres (HS) and $t = 0.8$ for Maxwellian molecules (MM). The reference solution is computed by the same method with a very small time step.

In Figure 13 we show the L_2 error for the fourth order moment of the distribution function f for Maxwellian molecules. In Figure 14 the L_2 error for the fourth order moment of f is reported in the case of hard sphere molecules. Observe that, in the case of Maxwellian molecules $\mu = 1$ while in the case of Hard Sphere particles μ is a constant upper bound for the collisional cross section ($\mu \gg 1$). This constraint implies that the L_2 norm of the error is larger, for equals choices of $\mu\Delta t/\varepsilon$, in the case of HS respect to the case of Maxwellian molecules.

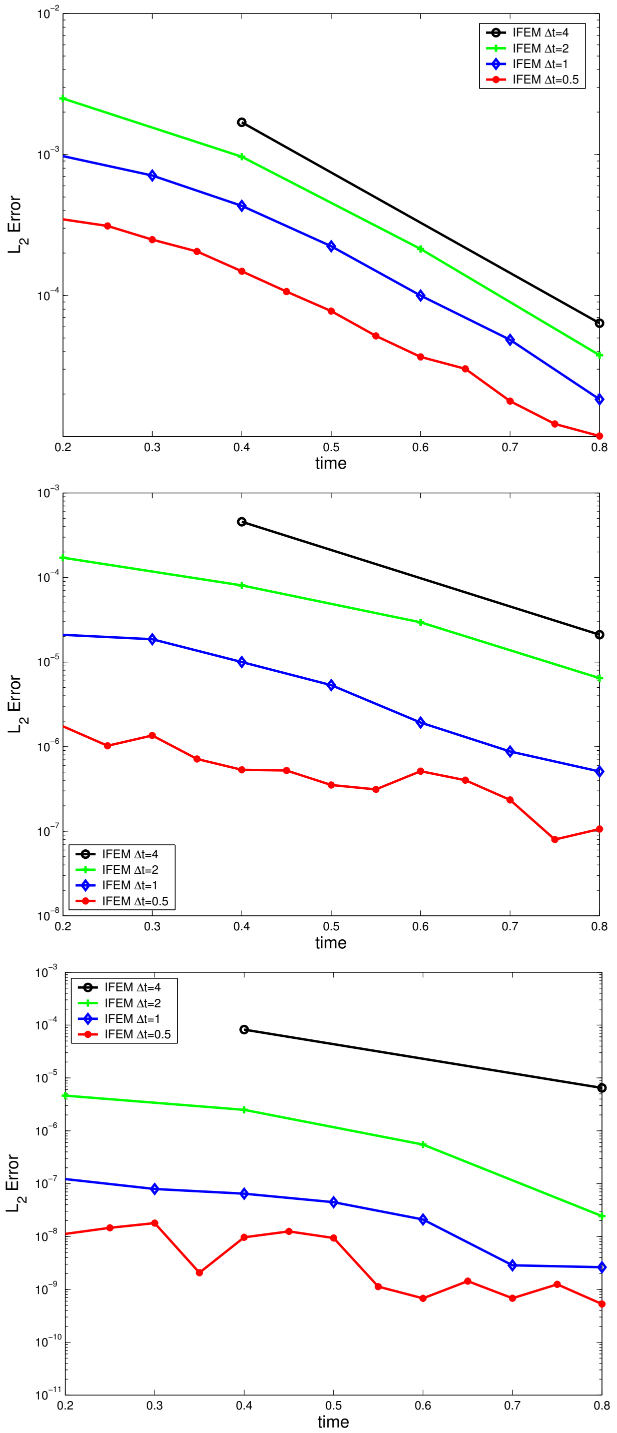


Figure 13: L_2 Error for the Fourth Moment Relaxation for the homogeneous relaxation problem with Maxwellian particles.

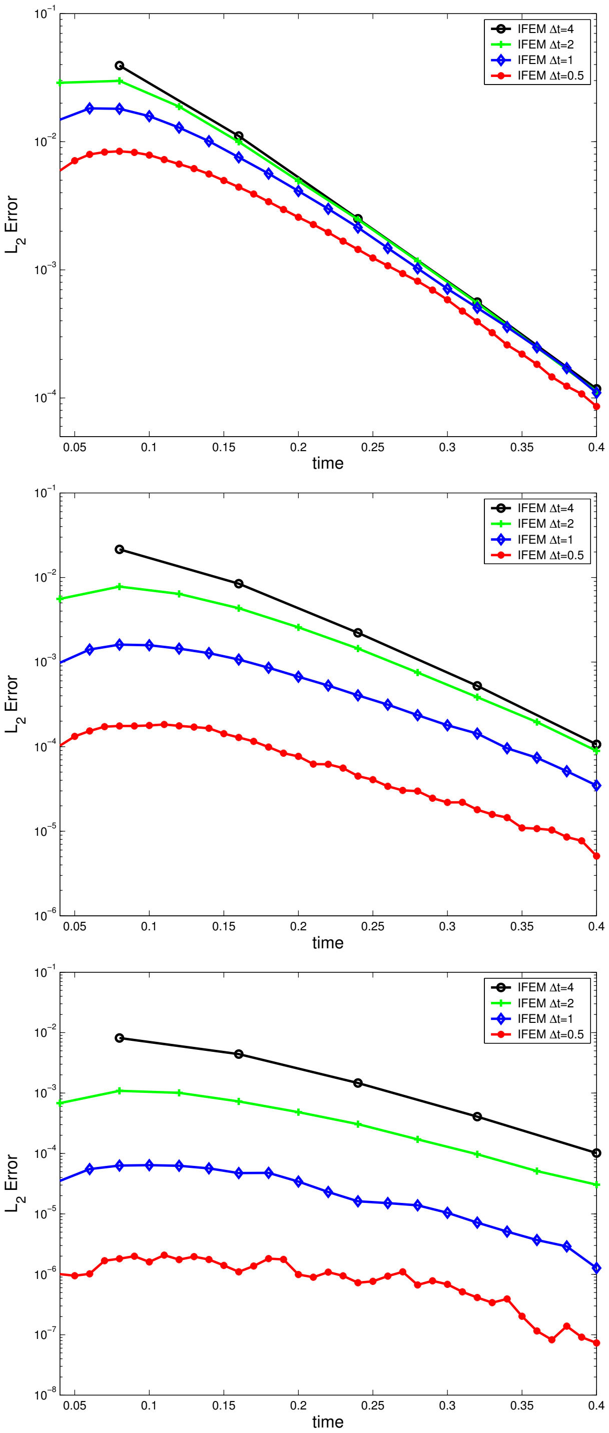


Figure 14: L_2 Error for the Fourth Moment Relaxation for the homogeneous relaxation problem with hard sphere particles.

Adaptivity

In this test we explore the possibility to use an adaptive value for μ in time. We use a first order TR scheme with adaptive time stepping and μ given by

$$\mu(t) = 3/2 \ \max_{v \in \Omega} \left(L(v) - \frac{Q_+}{f} \right),$$

where $\Omega = [-T/2, T/2]^3$. We use Maxwell molecules with the spectral method and 64^2 modes in 2D so that the velocity error can be neglected. We compare the results with a reference solution obtained with a 4th order Runge-Kutta with a fixed time step. The results are reported in Table 2.

The time evolution of μ is given in Figure 15. It shows that the method becomes more and more accurate as the solution approaches the Maxwellian state.

3.3.2 Non-homogeneous problems

Accuracy of the method

Here we consider the full non-homogeneous case where the collision operator is solved by spectral methods, the transport by the flux-positive schemes and the operator splitting is second order Strang splitting combined with the adaptive second order TR method. We test the overall accuracy of the method using as initial condition

$$f_0(x,v) = (1 + \beta \ \cos(k_0 \, x)) \ \exp(-v^2/2), \quad (x,v) \in [0,L] \times I\!R^2,$$

with periodic boundary conditions. The error is computed as

$$\varepsilon_{2\,h} = \max_{t \in (0,T)} (\|f_h(t) - f_{2\,h}(t)\|_1)/\|f_0\|_1,$$

and the results are given in Table 3. As expected the second order accuracy of the scheme is observed.

$\tau = \mu(t)\,\Delta t$	n_{Tot}	Numerical error $\varepsilon^{(1)}$ with a fixed $\Delta t = T/n_{Tot}$	Numerical error $\varepsilon^{(2)}$ with $\Delta t = \tau/\mu(t)$
0.010	50	0.0036	0.002
0.025	20	0.0080	0.007
0.050	11	0.0160	0.014
0.100	08	0.0300	0.025
0.200	07	0.0520	0.045
0.500	05	0.2000	0.090
1.000	03	XXXX	0.150
3.000	02	XXXX	0.040
5.000	01	XXXX	0.006

Table 2: Comparison between fixed Δt and $\Delta t = \tau/\mu(t)$

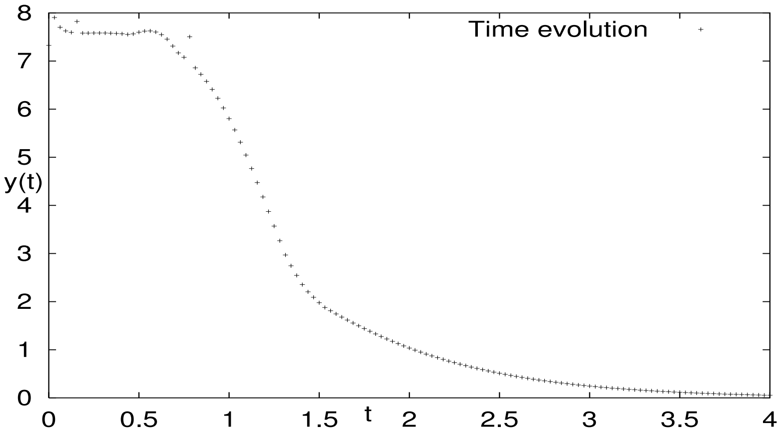


Figure 15: Time evolution of the value $\mu(t)$.

Numerical parameters	relative l^1 error norm	$\varepsilon_{2h}/\varepsilon_h$
$n_x = 032$, $n_{v_x} = n_{v_y} = 08$, $\Delta t = 0.100$	$\varepsilon_{4h} = 0.3835$	5.20
$n_x = 064$, $n_{v_x} = n_{v_y} = 16$, $\Delta t = 0.050$	$\varepsilon_{2h} = 0.0738$	4.35
$n_x = 128$, $n_{v_x} = n_{v_y} = 32$, $\Delta t = 0.025$	$\varepsilon_h = 0.0169$	X

Table 3: Convergence results

Stationary shock profile

We consider a stationary shock wave problem for the Boltzmann equation solved on a finite domain $-L < x < L$ with boundary conditions that the incoming flux of particles at $x = \pm L$ is distributed according to the Maxwellian flux $vM^{\pm}(v)$. As initial data, we take $f(x, v, 0) = M(\rho, u, T)$, with

$$\rho = 1.0, \quad T = 1.0, \quad \mathcal{M} = 2.0, \qquad L > x > 0,$$

where $\mathcal{M}$ is the Mach number. The mean velocity is

$$u_x = -\mathcal{M}\sqrt{\gamma T}, \quad u_y = 0,$$

with $\gamma = 5/3$.

The values for ρ, u and T for $x < 0$ are given by the Rankine-Hugoniot conditions (Whitham, 1974).

The profiles are shown in Figure 16 for different Knudsen numbers. As a reference solution we report also the solution obtained by Monte Carlo methods.

Riemann problem

This test deals with the numerical solution of the non homogeneous $1D \times 2D$ Boltzmann equation for hard sphere molecules ($\alpha = 1$). We have computed an approximation for different Knudsen numbers, from rarefied regime up to the fluid limit. The solution in the

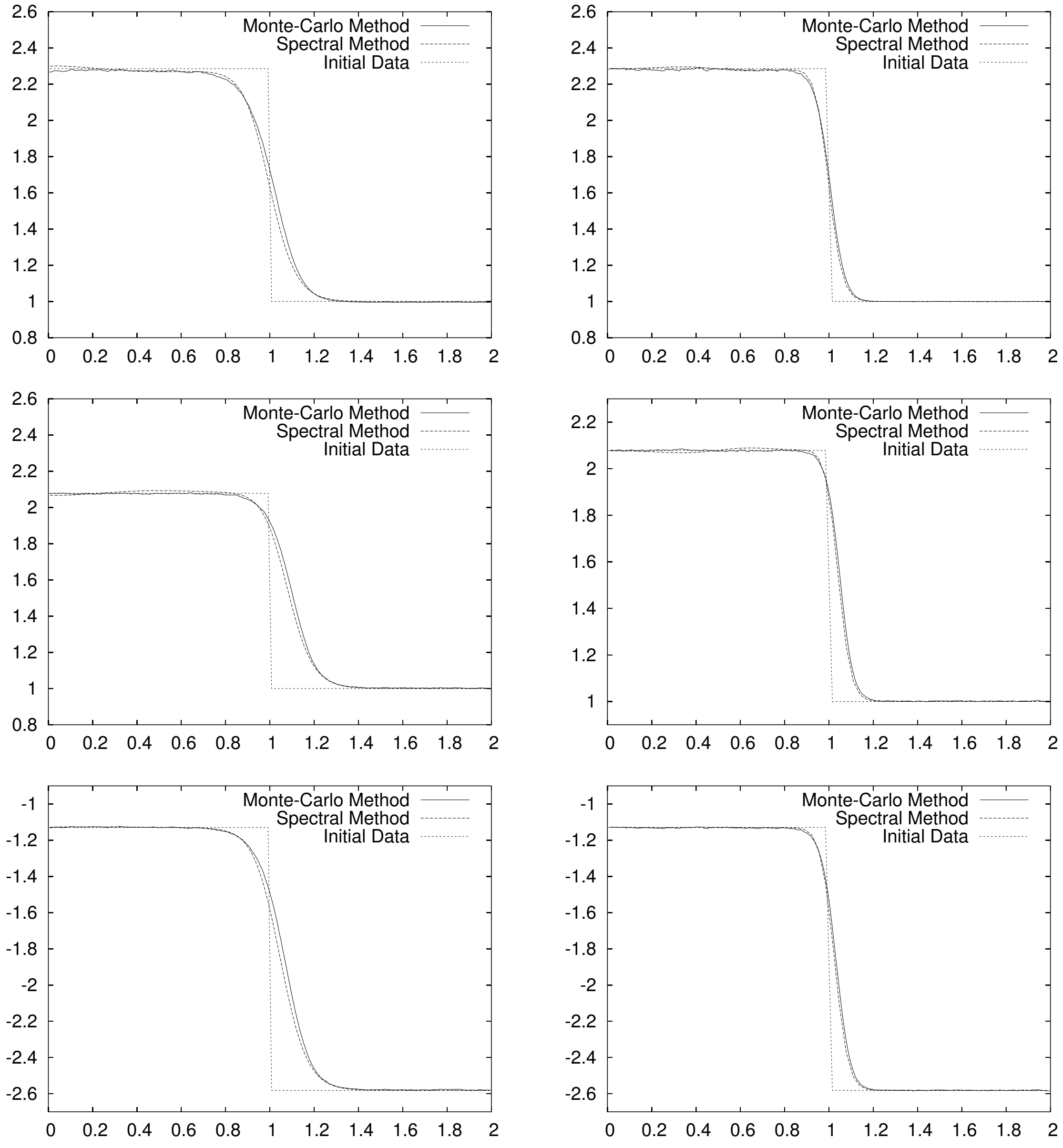


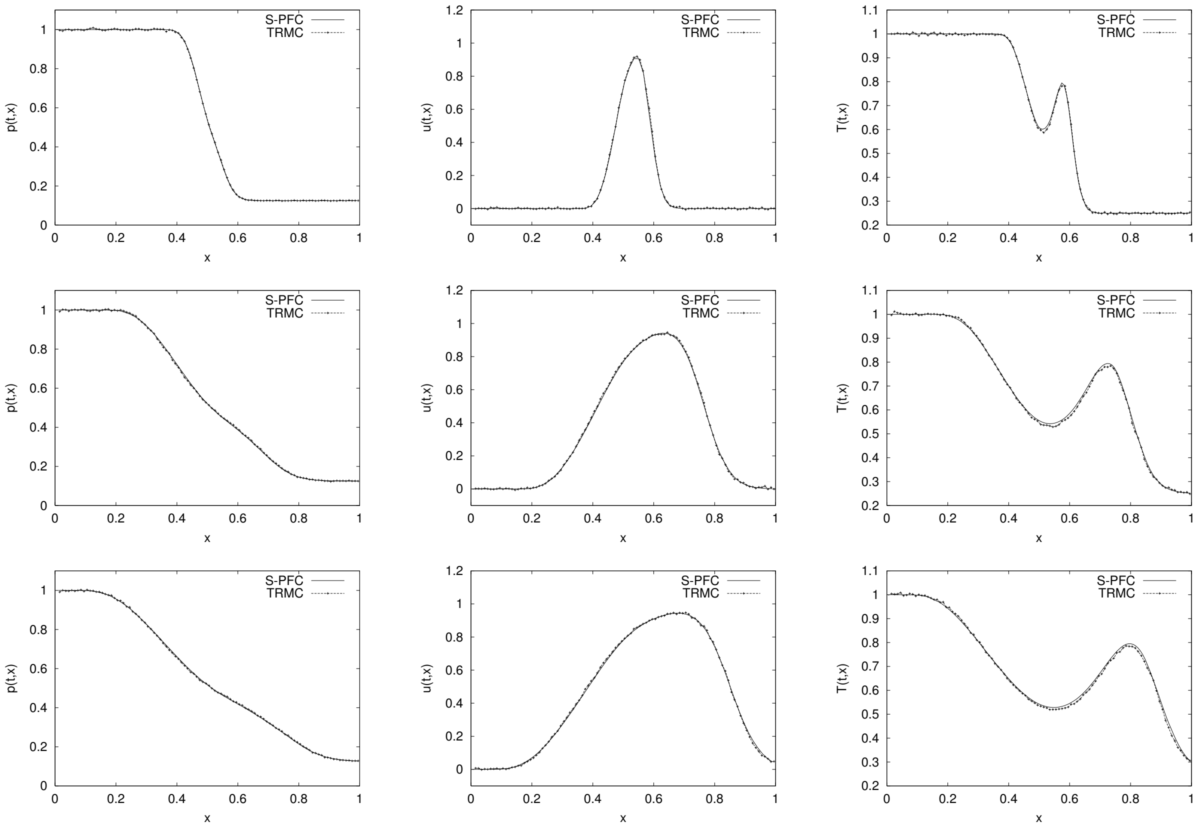
Figure 16: Shock profiles at Mach 2: $\varepsilon = 10^{-1}$ (left) and $\varepsilon = 0.05$ (right). From top to bottom: density ρ, mean velocity u and temperature T.

hydrodynamic limit is also compared with the numerical solution of Euler system. The initial data is given by

$$\begin{cases} (\rho_l, u_l, T_l) = (1, 0, 1) & \text{if } 0 \leq x \leq 0.5, \\ \\ (\rho_r, u_r, T_r) = (0.125, 0, 0.25) & \text{if } 0.5 < x \leq 1, \end{cases}$$

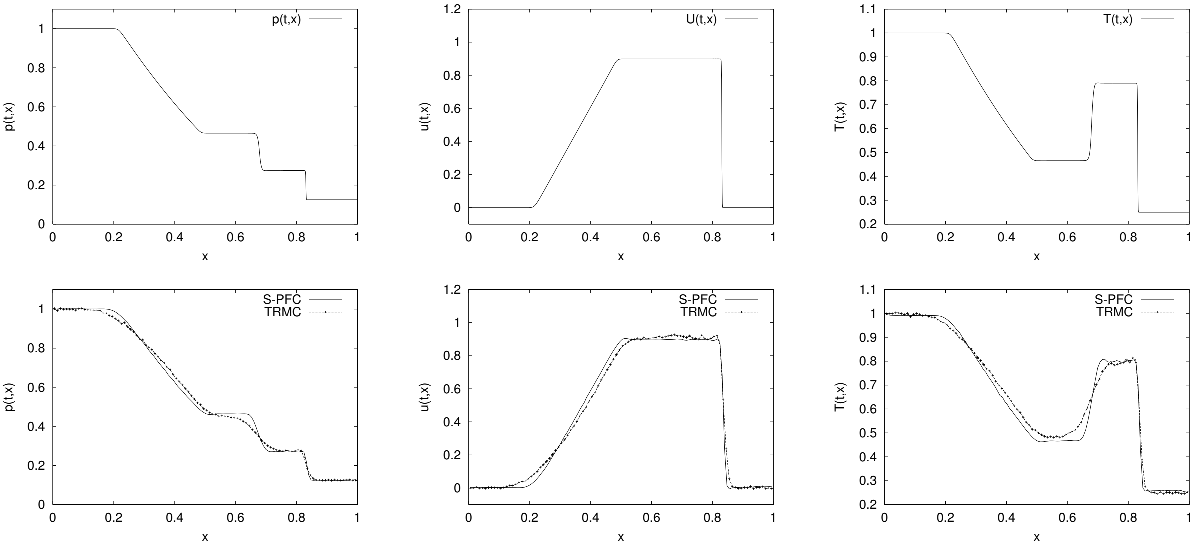
In Fig. 17 we plot the results obtained in the rarefied regime ($\varepsilon = 10^{-2}$) using the Spectral-PFC scheme and a Monte Carlo method (TRMC) as a comparison. The TRMC method is used with 100 cells in x containing 100 particles whereas the Spectral-PFC scheme is used with 64 points in x and the size of the velocity grid is 64×64 points for the transport and the total number of modes 32×32. We observe that the two solutions are in this case very comparable even if small oscillations, due to the statistical noise, persist. We also give the result of the computations close to the Euler limit ($\varepsilon = 10^{-4}$) using 128 space cells for the Spectral-PFC method.

Finally, the profiles obtained with TRMC and Spectral-PFC methods are reported in Fig. 18. On the opposite, using a small time step ($\Delta t = 0.001$), an accurate solution is obtained by the Spectral-PFC method, which is much less diffusive then the Monte Carlo methods.



(1) (2) (3)

Figure 17: Riemann problem ($k_n = 10^{-2}$): *evolution of (1) the density ρ, (2) mean velocity u and (3) temperature T at* $t = 0.05$, 0.15, 0.20.



(1) (2) (3)

Figure 18: Riemann problem ($k_n = 10^{-4}$): *(1) the density ρ, (2) mean velocity u and (3) temperature T at $t = 0.20$ obtained by the central scheme for Euler equations (up) and by Spectral-PFC and TRMC methods for Boltzmann equations.*

The ghost effect

Consider a gas between two plates at rest in a finite domain. In this situation, the stationary state at a uniform pressure (the velocity is equal to zero and the pressure is constant) is an obvious solution of the Navier-Stokes equations; the temperature field is determined by the heat conduction equation

$$u = 0, \quad T = C \quad - \nabla_x(T^{1/2}\nabla_x T) = 0.$$

On the other hand, if we move the plate by a velocity proportional to the Knudsen number, then the macroscopic fields (density and temperature profiles) will be affected by the flow, even for vanishing Knudsen number. This effect, called ”ghost effect”, is predicted by the Hilbert expansion of the Boltzmann equation in terms of the Knudsen number, and it is rather difficult to capture numerically, since the flow velocity is very small (Sone et al., 1996). The results show that the numerical solution agrees with one obtained by the asymptotic theory and not with the one obtained from the heat conduction equation; this result is a confirmation of the validity of the asymptotic theory.

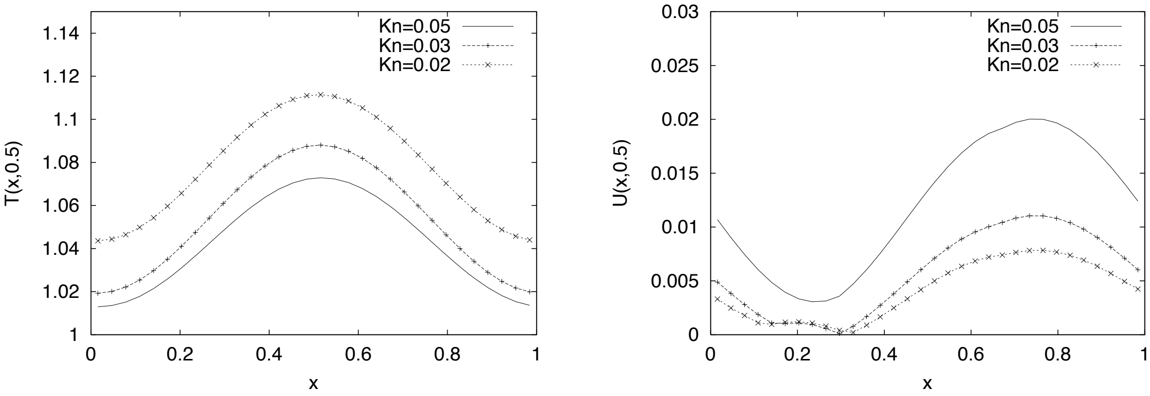
Thus consider two parallel plates, both with temperature distribution

$$T_w(x) = 1 - 0.5\,\cos(2\,\pi\,x); \quad \forall x \in (0,1),$$

in slow motion with velocity

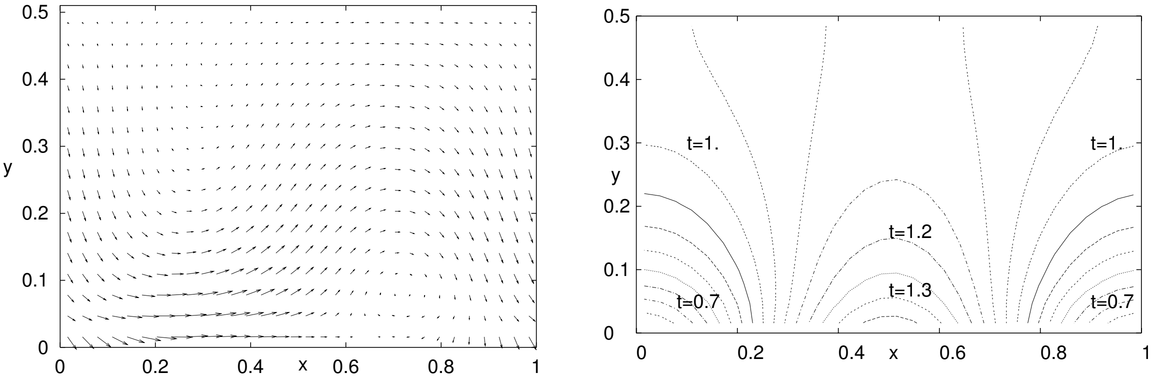
$$u_w(x) = (\varepsilon, 0).$$

We use the hard spheres model with diffusive b.c. on the walls and periodic in x. The cross section of temperature and velocity profile are shown in Figure 19 for various values of the Knudsen number, while velocity field and isothermal lines are reported in Figure 20, for Knudsen number $\varepsilon = 0.02$.



Temperature at $y = 0.5$ Mean velocity u at $y = 0.5$

Figure 19: Ghost effect: temperature and mean velocity along $y = const$ for various Knudsen numbers $\varepsilon = 0.05, 0.02, 0.01$.



velocity field u isothermal lines

Figure 20: Ghost effect: velocity field, and isothermal lines; Knudsen number $\varepsilon = 0.02$.

4 Other asymptotic preserving methods

In this section we shall consider alternative approaches to splitting methods when dealing with the time integration of kinetic equations with stiff collision operators. This could be an advantage for the construction of high-order or well-balanced schemes, i.e. schemes that preserve the stationary solutions. First we discuss semilagrangian schemes and then IMEX schemes for BGK like models. As we shall see, the BGK model allows a simple implicit treatment. This is useful *per se*, and it provides a building block for effective treatment of the Boltzmann equation near the fluid dynamic limit.

4.1 Semilagrangian methods for the BGK model

Here we will focus on implicit semilagrangian scheme for the numerical solution of the BGK model of the Boltzmann equation. We shall restrict to the BGK equation in one space dimension. More details on the method can be found in Santagati (2007).

4.1.1 A basic first order scheme

Let us rewrite the BGK model in one dimension:

$$\frac{\partial f}{\partial t} + v\frac{\partial f}{\partial x} = \frac{1}{\varepsilon}(M[f] - f). \tag{86}$$

The numerical scheme for the solution of Eq. (86) is based on the characteristic formulation of the problem (86),

$$\begin{aligned} &\frac{df}{dt} = \frac{1}{\varepsilon}(M\,[f] - f),\\ &\frac{dx}{dt} = v,\\ &x(0) = \tilde{x}, \quad f(0,x,v) = f_0(\tilde{x},v) \quad t \geq 0, \quad x, v \in \mathbb{R}. \end{aligned} \tag{87}$$

For simplicity we assume constant time step Δt and uniform grid in physical and velocity space, with mesh spacing Δx and Δv respectively, and denote the grid points by $t^n = n\Delta t$, $x_i = i\Delta x$, $i = 1, \dots, N_x$, $v_j = j\Delta v$, $j = -N_v, \dots, N_v$, where N_x and $2N_v + 1$ are the numbers of grid nodes in space and velocity, respectively. Let f_{ij}^n denote the approximate solution of the problem (87) at time t^n in each spatial and velocity node.

We start by considering first order accurate schemes.

An explicit first order semilagrangian scheme could be constructed by computing an approximation $\tilde{f}$ of $f(t^{n+1}, x_i + v_j\Delta t, v_j)$ as

$$\tilde{f}(t^{n+1}, x_i + v_j\Delta t, v_j) = f_{ij}^n + \frac{\Delta t}{\varepsilon}(M_{ij}^n - f_{ij}^n) \tag{88}$$

The function $\tilde{f}$ computed in this way at the new time step does not lie on a grid. The values of f_{ij}^{n+1} could be reconstructed from the computed values $\tilde{f}$ by a suitable interpolation back on the grid points (see Figure 21). Let us denote by $\tilde{x}_i = x_i + v_j\Delta t$, and let us assume that this point is between point x_k and x_{k+1}. Then the function f_{kj}^{n+1} can be

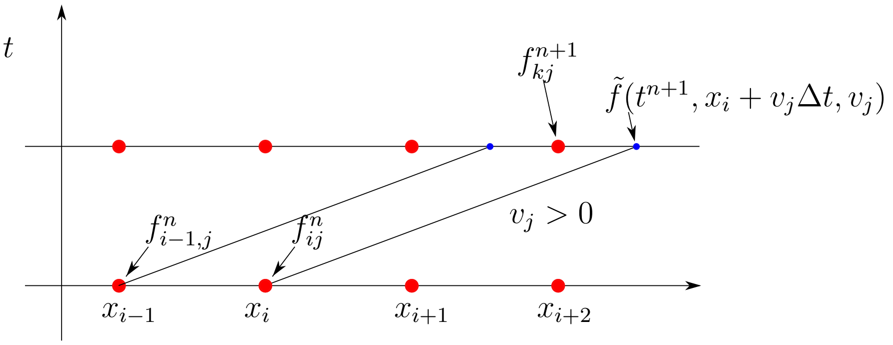


Figure 21: Propagation of the information along characteristics in the explicit scheme. The computed point does not lie on the grid, and some interpolation is needed in order to compute f_{kj}^{n+1}. In this case $k = i + 2$.

reconstructed by simple linear interpolation using the computed value at points $\tilde{x}_i$ and $\tilde{x}_{i-1}$.

The Maxwellian $M_{ij}^n = M(v_j, \{\rho_i^n, u_i^n, T_i^n\})$ is computed as follow

$$M_{ij}^n = \frac{\rho_i}{(2\pi R\tilde{T}_i)^{1/2}} \exp\left(-\frac{|v_j - u_i|^2}{2RT_i}\right). \tag{89}$$

This formula requires the computation of the discrete moments of $\{f_{ij}^n\}$. This can be done by using a numerical approximation of the integrals computed in (21). Following the notation in Mieussens (2000), the discrete velocity grid may be denoted by $\mathcal{V}$, which is composed of $2N_v + 1$ nodes, and the moments of any quantity g, $< g >$, can be approximated by a quadrature rule on $\mathcal{V}$. Let $< g >_{\mathcal{V}}$ denote the approximation of $< g >$, where $\mathcal{V}$ is the set of $2N_v + 1$ indices matching the velocity grid nodes. By this way we compute the moments of the Maxwellian at each grid nodes $\{x_i\}$,

$$(\rho_i, \rho_i u_i, E_i) =< f_i^n \phi(v) >_{\mathcal{V}}$$

As quadrature rule we use summation over $\mathcal{V}$ times Δv, providing spectral accuracy for smooth functions on compact support. The grid $\mathcal{V}$ is chosen to include most of the mass. For a given number of nodes N_v, an optimal choice of the grid is obtained as a compromise between the extension of the velocity domain and the resolution of the grid.

Once the moments are computed on the grid, they can be in turn computed in $\tilde{x}_i$, by a suitable interpolation formula, so that the Maxwellian gets easily evaluated. Details about the WENO interpolation can be found in Cockburn et al. (1998), or in Russo and Santagati (2011).

The scheme (88) can be used to perform the time step. The scheme is first order accurate. A more sophisticated time integrator, coupled with a more accurate interpolation, would provide greater accuracy in time. Notice that, because of the semilagrangian nature of the method, there is no CFL-type stability restriction on the time step due to convection. However, such scheme would suffer from stability restriction on the time step due to the collision term when the collision time ε is small.

In order to circumvent the stiffness arising from when ε is small, it is possible to resort to an implicit formulation.

By applying simple implicit Euler on the characteristic equation in order to compute f_{ij}^{n+1} one obtains:

$$f_{ij}^{n+1} = \tilde{f}_{ij}^{n} + \frac{\Delta t}{\varepsilon}(M_{ij}^{n+1} - f_{ij}^{n+1}) \tag{90}$$

The quantity $\tilde{f}_{ij}^{n} \approx f(t^n, x_i - v_j\Delta t, v_j)$ can be computed by suitable reconstruction from $\{f_{\cdot j}^{n}\}$; linear reconstruction will be sufficient for first order scheme, while higher order reconstruction, such as ENO or WENO, could be used to provide higher order non oscillatory reconstruction.

The equation cannot be immediately solved for f_{ij}^{n+1}, because the Maxwellian depends from f^{n+1} itself. However, one can act as follows: let us take the moments of both sides of Eq. (90). This is obtained at a discrete level multiplying both sides by $\phi_j\Delta v$, where $\phi_j = \{1, v_j, |v_j|^2\}$, and summing over j. We denote this procedure by $< \cdot >_{\mathcal{V}}$, i.e. for any quantity h_j, we define

$$< h >_{\mathcal{V}} \equiv \sum_{v_j\in\mathcal{V}} h_j\Delta v$$

Then we have

$$< (f_{ij}^{n+1} - \tilde{f}_{ij}^{n})\phi_j >_{\mathcal{V}} = \frac{\Delta t}{\varepsilon} < (M_{ij}^{n+1} - f_{ij}^{n+1})\phi_j >_{\mathcal{V}} \tag{91}$$

The right hand side is zero, because, by definition, the Maxwellian at time t^{n+1} has the same moments as f^{n+1}. As a consequence, the moments of f^{n+1} can be computed as moments of $\tilde{f}_{ij}$. More explicitly, one has

$$\begin{aligned} \rho_i^{n+1} &= \sum_{v_j\in\mathcal{V}} \tilde{f}_{ij}^{n}\,\Delta v \\ u_i^{n+1} &= \frac{1}{\rho_i^{n+1}}\sum_{v_j\in\mathcal{V}} \tilde{f}_{ij}^{n} v_j\,\Delta v \\ T_i^{n+1} &= \frac{1}{dR\rho_i^{n+1}}\sum_{v_j\in\mathcal{V}} \tilde{f}_{ij}^{n}(v_j - u_i)^2\,\Delta v \end{aligned} \tag{92}$$

where d denotes the dimension of velocity space (in our case $d = 1$).

Once the density, mean velocity and temperature are computed, then the Maxwellian at the new time step can be explicitly computed:

$$M_{ij}^{n+1} = M(v_j; \{\rho_i^{n+1}, u_i^{n+1}, T_i^{n+1}\}).$$

Once the Maxwellian is known, the distribution function can be explicitly computed:

$$f_{ij}^{n+1} = \frac{\varepsilon\tilde{f}_{ij}^{n} + \Delta t M_{ij}^{n+1}}{\varepsilon + \Delta t}. \tag{93}$$

Notice that since we use a semilagrangian scheme, the geometrical CFL condition is always satisfied, since the value $\tilde{f}_{ij}^{n}$ is computed by interpolation from points that surround points $\tilde{x}_{ij} = x_i - v_j\Delta t$. For example, in the case illustrated in Figure 22, $x_{ij} \in [x_k, x_{k+1}]$, with $k = i - 2$, and for the first order scheme the value of $\tilde{f}_{ij}^{n}$ can be obtained by linear interpolation from the values f_{kj}^{n} and $f_{k+1,j}^{n}$.

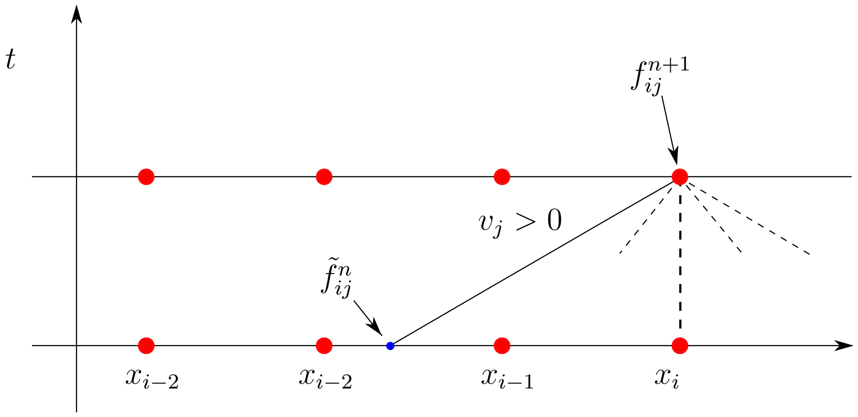


Figure 22: Propagation of the information along characteristics in the implicit scheme. The foot of the characteristics does not lie on the grid, and some interpolation is needed in order to compute $\tilde{f}_{ij}^{n}$

Remark 8 *Since the moments are computed by a quadrature formula, it is not properly true that, in the discrete formulation, $M[f]$ and f have the same moments. To get an insight on this aspect see Mieussens (2000). In that paper the author introduces the notion of a discrete Maxwellian, which is more consistent with the discrete formulation of the problem. The discrete BGK model obtained using such Maxwellian is conservative and entropic. By enough large number of grid points in velocity, the continuous and discrete Maxwellians give comparable results. However, for coarse discretization in velocity, the discrete Maxwellian introduced in Mieussens (2000) produces better results.*

In next section we shall show how to generalize the procedure and construct higher order schemes.

4.1.2 High order time discretization

System (87) is a typical ordinary differential equation with relaxation, to be solved in the characteristics framework. Relaxation time lies in a very wide range. It typically extends from order one to very small values compared to the time scale of the problem. For this reason, we treat the relaxation operator by L-stable diagonally implict Runge Kutta (DIRK) schemes (Hairer and Wanner, 1996; Pareschi and Russo, 2000a, 2005), which provides enough stability to ensure the AP property (see Sec. 1.7). DIRK schemes are defined by the triangular $\nu\times\nu$ matrix, $A = (a_{lk})$, and the coefficient vectors, $c = (1, ..., c_\nu)^T$ and $b = (1, ..., \nu)^T$, which are derived by imposing accuracy and stability constraints. They characterize completely a DIRK scheme, which can be rappresented by the Butcher's *tableaux*

$$\begin{array}{c|c} c & A \\ \hline & w^T \end{array}$$

The internal stages are practically evaluated by a sequence of elementary implicit Euler steps, because, at each stage, only the last stage value is unknown, due to the triangular structure of matrix A. Scheme (90) corresponds to implicit Euler scheme. It will be denoted $M1$. The DIRK methods considered in this work are

$$M2 = \begin{array}{c|cc} \alpha & \alpha & \\ 1 & 1-\alpha & \alpha \\ \hline & 1-\alpha & \alpha \end{array}, \qquad M3 = \begin{array}{c|ccc} 1/2 & \gamma & & \\ (1+\gamma)/2 & (1-\gamma)/2 & \gamma & \\ 1 & 1-\delta-\gamma & \delta & \gamma \\ \hline & 1-\delta-\gamma & \delta & \gamma \end{array}$$

which are a second and a third order L-implicit schemes Pareschi and Russo (2000a). The coefficients are

$$\alpha = 1 - \frac{\sqrt{2}}{2}, \quad \gamma = 0.4358665215, \quad \delta = -0.644373171.$$

Some DIRK schemes have the property that the last row of matrix A equals the vector of the weights b. Such schemes are called *stiffly accurate*. This property is related to the L-stability of the scheme. For more details see, for example, the classical book by Hairer and Wanner (1996).

Here we apply the Runge-Kutta scheme along the characteristics. The numerical solution is obtained as

$$f_{ij}^{n+1} = \tilde{f}_{ij} + \Delta t \sum_{\ell=1}^{\nu} b_\ell \tilde{K}_{ij}^{(\ell)} \tag{94}$$

The quantity

$$\tilde{K}_{ij}^{(\ell)} = \frac{1}{\varepsilon}(M[\tilde{F}_{ij}^{(\ell)}] - \tilde{F}_{ij}^{(\ell)})$$

are the so called Runge-Kutta fluxes, and have to be evaluated along the characteristics, and depend on the so called stage values $\tilde{F}_{ij}^{(\ell)}$. In a standard DIRK methods, the ℓ-th stage value, say $\tilde{F}_{ij}^{(\ell)}$, is evaluated by solving an implicit equation involving only $\tilde{F}_{ij}^{(\ell)}$, since the previous stage values have already been computed, due to the triangular structure of the matrix A. In our case, however, this is not so easy, because if the point corresponding to stage ℓ along the characteristics is not a grid point, it is not possible to compute the moments of the Maxwellian at that point in space-time. For this reason, one has to resort to a triangular scheme, which is illustrated, for a two stage scheme, with the help of Figure 23.

Two kinds of stage values will be needed: the stage values along the characteristics, denoted by $\tilde{F}_{ij}^{(\ell)}$, which are needed for the computation of the numerical solution, and the stage values on the grid, denoted by $\hat{F}_{ij}^{(\ell)}$, which can be computed implicitly.

The two-stage stiffly-accurate scheme $M2$ works as follows.

First we compute $\hat{F}_{ij}^{(1)}$ by

$$\hat{F}_{ij}^{(1)} = \frac{\varepsilon \hat{f}_{ij}^{(1)} + \Delta t \hat{M}_{ij}^{(1)}}{\varepsilon + \Delta t} \tag{95}$$

Here the Maxwellian $\hat{M}_{ij}^{(1)} = M[\hat{F}_{ij}^{(1)}]$ can be computed by computing the moments of $\hat{f}_{ij}^{(1)} = f_j^n(x_i - \alpha v_j \Delta t)$ by suitable space reconstruction at time t^n.

Once the implicit step is solved, the value of the Runge-Kutta flux $\hat{K}_{ij}^{(1)}$ can be easily computed at the grid points $(x_i, t^n + \alpha\Delta t)$. Then the fluxes are computed by high order interpolation on the intermediate nodes along the characteristics (marked by a blue

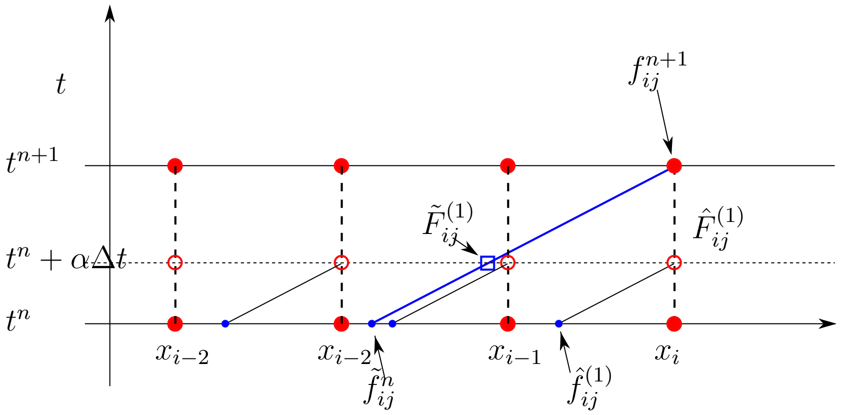


Figure 23: Construction of two-stage stiffly-accurate DIRK. The first stage is computed on the grid points $(t^n+\alpha\Delta t, x_i)$ (empty red circle) by an implicit step. The RK fluxes are computed on the same points, and then interpolated to points $(t^n+\alpha\Delta t, x_i-(1-\alpha)v_j\Delta t)$, denoted by a small blue square. Finally, the second stage value along the characteristics is computed on the grid. Such stage value corresponds to the numerical solution implicit scheme

square). Once such values, $\tilde{K}_{ij}^{(1)}$, are computed, then the value of the numerical solution can be computed by

$$\tilde{F}_{ij}^{(2)} = \tilde{f}_{ij}^n + \Delta t\left(a_{21}\tilde{K}_{ij}^{(1)} + a_{22}\frac{1}{\varepsilon}(M[\tilde{F}_{ij}^{(2)}] - \tilde{F}_{ij}^{(2)})\right) \tag{96}$$

Notice that in this $f_{ij}^{n+1} = \tilde{F}_{ij}^{(2)}$, because the scheme is stiffly accurate, i.e. $a_{21} = b_1$ and $a_{22} = b_2$.

The generalization of the procedure to high order schemes is straightforward, and can be found, for example in Russo and Santagati (2011).

Remark 9 *In practice the Runge Kutta fluxes can be computed from the internal stages. For example*

$$\frac{\Delta t}{\varepsilon}(M_{ij}^{(1)} - F_{ij}^{(1)}) = \frac{F_{ij}^{(1)} - \tilde{f}_{ij}^n}{a_{11}}.$$

Hence the scheme can be used in the limit $\varepsilon \to 0$, with no constraint on the time step amplitude.

4.2 Numerical tests

These tests are aimed to verify the accuracy (test 1) and the shock capturing properties (test 2) of the schemes.

4.2.1 Regular velocity perturbation

This test has been proposed in Pieraccini and Puppo (2007). The solution is smooth, and the accuracy can be tested. Initial velocity profile is given by

$$u_0(x) = \frac{1}{\sigma}\left(\exp\left(-(\sigma x - 1)^2\right) - 2\exp\left(-(\sigma x + 3)^2\right)\right), \quad x \in [-1, 1]$$

where σ is a positive constant parameter. Initial density and temperature profiles are uniform, with constant value, $\rho = 1$ and $T = 1$ respectively. The initial condition for the distribution function is the Maxwellian, computed by given macroscopic fields. The boundary conditions are imposed assuming that beyond the computational domain the distribution function is a Maxwellian with prescribed moments. Two regimes (rarefied and fluid) have been investigated, corresponding to different Knundsen numbers, $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-6}$. The final time, for both cases was 0.04, large enough to reach thermodynamic equilibrium. Accuracy and conservation tests have been performed at the final time. The errors has been computed using a reference solution, defined on a finer grid, with $N_x = 1280$ and $N_v = 20$. The test case has been performed using $N_v = 20$ (as for the reference solution),for each spatial grid nodes number, uniformly spaced in [-10,10].

The relative errors and order of accuracy are shown in Tables 4-7, for the schemes $M2$ and $M3$. Notice that only a moderate improvements is obtained using $M3$ in place of $M2$. The reason is the following. The space reconstruction in both schemes is WENO 2-3, which guarantees third order accuracy for smooth functions. Since space errors are dominant with respect to time errors, then only a small improvement is gained by a more accurate time integrator. This also explains why for small number of grid points the order or accuracy appears between two and three.

Also, notice that for smaller values of ε a strong degradation of the accuracy is observed. In fact, in the fluid dynamic regime, only first order accuracy in time is guaranteed by this approach.

Conservation errors have been investigated. Despite the schemes are not strictly conservative, conservation properties are maintained with good accuracy, even for a moderate number of grid points.

4.2.2 Riemann problem

This test allows us to evaluate the capability of our class of schemes in capturing shocks, contact discontinuities and the density profile in a rarefaction. The macroscopic fields are initially assigned in the domain, satisfying the Rankine-Hugoniot shock jump conditions. In particular we are interested in the behavior in the fluid regime. Here we illustrate the results in the moments, i.e. density, velocity and temperature profiles, for $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-6}$, respectively. As in test 1, the boundary conditions are imposed by Maxwellians computed by prescribed macroscopic moments.

For this test two values ε are employed , $\varepsilon = 10^{-2}$ and $\varepsilon = 10^{-6}$. $N_v = 60$ nodes are used in the range [-10,10] of the discrete velocity domain, as in Pieraccini and Puppo (2007).

As it appears from the results, the scheme is able to capture the fluid dynamic limit for very small values of the relaxation time, where the evolution of the moments is governed by the underlying Euler equations.

Remark 10 *The scheme presented here is not in conservation form. It is possible to construct a conservative version of the scheme, by using the non conservative scheme as a predictor, and a postprocessing that acts as a conservative corrector to be used at each time step. The details of the scheme will be found in Russo and Santagati (2011).*

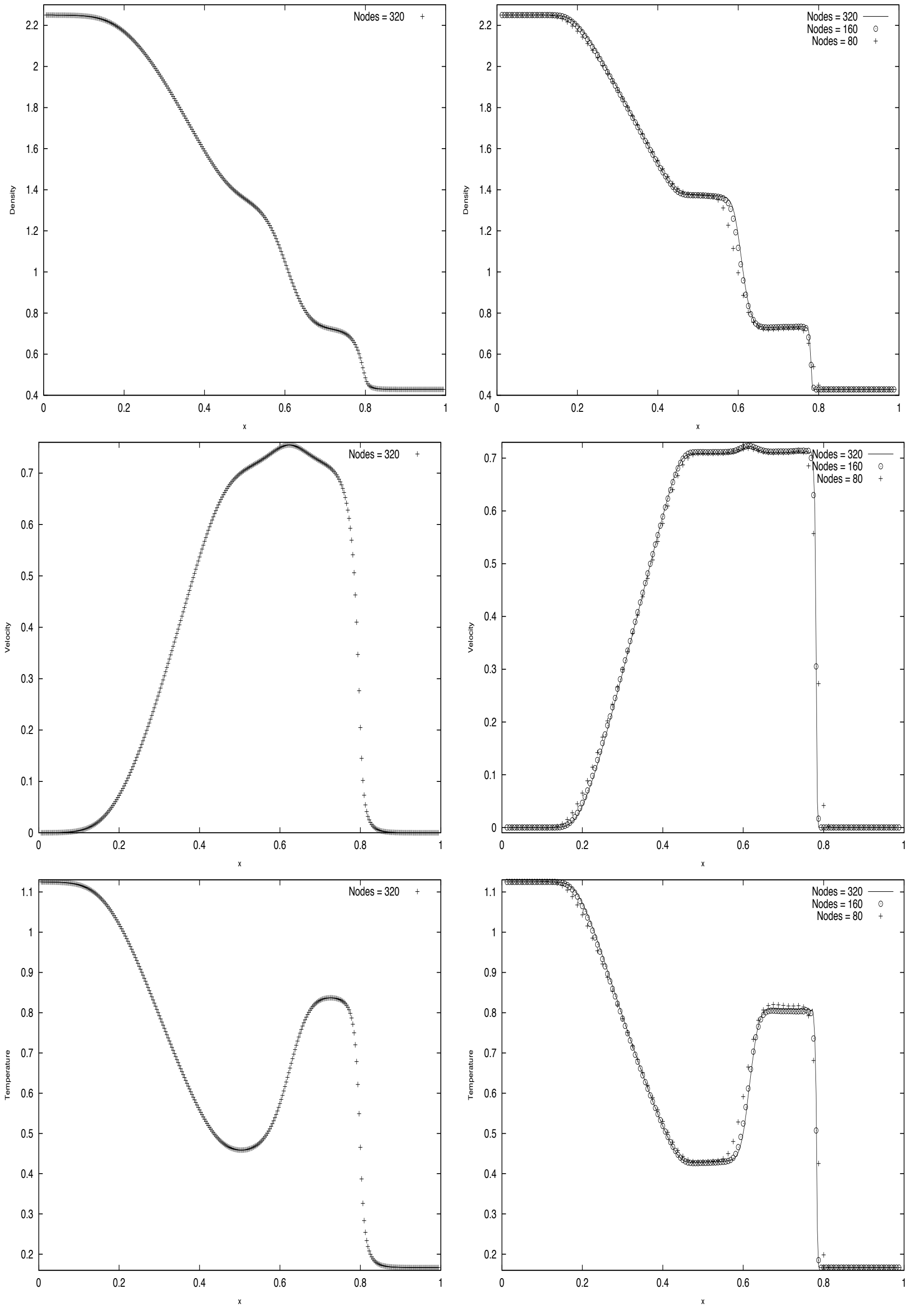


Figure 24: Riemann problem. From the top to the bottom, density, velocity and temperature. Left column $\varepsilon = 10^{-2}$. Right column $\varepsilon = 10^{-6}$. CFL=9.44.

L^2-Relative errors				L^2-Orders		
N_x	Density	Velocity	Temperature	Dens.	Vel.	Temp
20	2.54838e-03	2.35049e-03	4.63423e-03	-	-	-
40	5.57339e-04	3.64146e-04	9.17053e-04	2.193	2.690	2.337
80	8.41532e-05	5.21314e-05	1.28062e-04	2.727	2.804	2.840
160	1.17817e-05	8.94658e-06	2.53109e-05	2.836	2.543	2.339
320	1.69746e-06	1.95126e-06	6.47027e-06	2.795	2.197	1.968

Table 4: Scheme M2, $\varepsilon = 10^{-2}, CFL = 4.5$.

L^2-Relative errors				L^2-Orders		
N_x	Density	Velocity	Temperature	Dens.	Vel.	Temp.
20	2.96809e-03	3.15227e-03	6.37101e-03	-	-	-
40	6.57722e-04	6.05216e-04	1.94043e-03	2.174	2.381	1.715
80	1.11120e-04	1.36059e-04	5.26168e-04	2.565	2.153	1.883
160	2.25137e-05	4.61239e-05	1.59816e-04	2.303	1.561	1.719
320	6.10643e-06	1.63245e-05	5.05549e-05	1.882	1.498	1.660

Table 5: Scheme M2, $\varepsilon = 10^{-6}, CFL = 4.5$.

L^2-Relative errors				L^2-Orders		
N_x	Density	Velocity	Temperature	Dens.	Vel.	Temp.
20	2.41539e-03	1.97185e-03	4.36445e-03	-	-	-
40	4.93444e-04	2.90747e-04	8.14397e-04	2.291	2.762	2.422
80	7.36995e-05	4.27296e-05	1.14397e-04	2.743	2.766	2.832
160	1.06248e-05	5.91413e-06	1.54660e-05	2.794	2.853	2.887
320	1.55051e-06	9.55269e-07	2.89414e-06	2.777	2.630	2.418

Table 6: Scheme M3, $\varepsilon = 10^{-2}, CFL = 4.5$.

L^2-Relative errors				L^2-Orders		
N_x	Density	Velocity	Temperature	Dens.	Vel.	Temp.
20	2.02024e-03	2.72023e-03	4.79517e-03	-	-	-
40	5.04007e-04	7.57946e-04	2.06197e-03	2.003	1.844	1.218
80	9.91878e-05	2.63921e-04	9.43964e-04	2.345	1.522	1.127
160	4.10972e-05	1.48278e-04	5.14779e-04	1.271	0.932	0.975
320	2.17256e-05	7.52320e-05	2.39514e-04	1.120	1.079	1.104

Table 7: Scheme M3, $\varepsilon = 10^{-6}, CFL = 4.5$.

4.3 IMEX Runge-Kutta schemes

The possibility of implementing a very efficient implicit solver for the BGK model, coupled with the fact that the BGK operator itself is a crude approximation of the Boltzmann

ε =1e-2			
N_x	Density	Momentum	Energy
20	3.19103e-04	6.45517e-04	6.44489e-04
40	8.68364e-05	2.43064e-05	1.53479e-04
80	1.86619e-05	6.93736e-06	3.38482e-05
160	2.21369e-06	6.03659e-07	3.82991e-06
320	2.47089e-07	6.48153e-08	4.23732e-07
640	2.75503e-08	5.92434e-09	4.57648e-08
1280	3.21756e-09	6.82768e-10	5.29436e-09
ε =1e-6			
N_x	Density	Momentum	Energy
20	3.86571e-04	8.49545e-04	8.59503e-04
40	1.25170e-04	4.22174e-05	2.34721e-04
80	3.17682e-05	1.53056e-05	6.50680e-05
160	4.65888e-06	1.86177e-06	9.55101e-06
320	5.94587e-07	2.25448e-07	1.20946e-06
640	7.47884e-08	2.69514e-08	1.52337e-07
1280	9.24536e-09	3.27055e-09	1.87483e-08

Table 8: Errors in conservation. Scheme M2.

collision operator, can be used as a tool to construct very effective solvers for the full Boltzmann equation. In this section we will consider this possibility in the context of Implicit-Explicit Runge-Kutta schemes.

4.3.1 General formulation of IMEX schemes

To this goal we introduce the general formulation of the IMEX schemes (Pareschi and Russo, 2005; Pieraccini and Puppo, 2007) for the BGK model. A general IMEX schemes applied to a kinetic equation of the form

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}(M[f] - f) \tag{97}$$

reads

$$F^{(i)} = f^n - \Delta t \sum_{j=1}^{i-1} \widetilde{a}_{ij} v \cdot \nabla_x F^{(j)} + \Delta t \sum_{j=1}^{\nu} a_{ij} \frac{1}{\varepsilon}(M[F^{(j)}] - F^{(j)}) \tag{98}$$

$$f^{n+1} = f^n - \Delta t \sum_{i=1}^{\nu} \widetilde{\omega}_i v \cdot \nabla_x F^{(i)} + \Delta t \sum_{j=1}^{\nu} \omega_i \frac{1}{\varepsilon}(M[F^{(i)}] - F^{(i)}). \tag{99}$$

The matrices $\widetilde{A} = (\widetilde{a}_{ij})$, $\widetilde{a}_{ij} = 0$ for $j \geq i$ and $A = (a_{ij})$ are $\nu \times \nu$ matrices such that the resulting scheme is explicit in $\nabla_x f$, and implicit in $M[f] - f$. In general, an IMEX Runge-Kutta scheme, is characterized by the above defined two matrices and the coefficient vectors $\widetilde{w} = (\widetilde{w}_1, .., \widetilde{w}_\nu)^T$, $w = (w_1, .., w_\nu)^T$. Since we want simplicity and efficiency in solving the algebraic equations corresponding to the implicit part we will consider only diagonally implicit Runge-Kutta (DIRK) schemes for the source terms ($a_{ij} = 0$, for $j > i$). The use of a DIRK scheme is enough to assure that the operator $\nabla_x f$ is always evaluated explicitly. The type of schemes described can be represented with a compact notation by a double Butcher tableau

$$\begin{array}{c|c} \widetilde{c} & \widetilde{A} \\ \hline & \\ & \widetilde{\omega}^T \end{array} \qquad\qquad \begin{array}{c|c} c & A \\ \hline & \\ & \omega^T \end{array}$$

where the coefficients $\widetilde{c}$ and c are given by the usual relation

$$\widetilde{c}_i = \sum_{j=1}^{i-1} \widetilde{a}_{ij}, \qquad c_i = \sum_{j=1}^{i} a_{ij}. \tag{100}$$

We refer to Pareschi and Russo (2005) for a discussion of the stability requirements and the derivation of the order condition of IMEX schemes up to third order. Let us remark that order conditions are simplified under the assumptions $\widetilde{c} = c$ and $\widetilde{w} = w$. The first assumption however prevents the scheme from being asymptotic preserving unless the initial data is well-prepared, namely consistent with the limiting equilibrium system.

As shown in Pareschi and Russo (2005) it is easy to prove the following:

Lemma 4 *If all diagonal element of the triangular coefficient matrix A that characterize the DIRK scheme are non zero, then*

$$\lim_{\varepsilon\to 0} F^{(i)} = M[F^{(i)}]. \tag{101}$$

Now if we multiply the IMEX method by the collision invariants $\phi(v) = 1, v, v^2$ and integrate the result in velocity space we obtain

$$\int_{\mathbb{R}^3} F^{(i)}\phi(v)\,dv = \int_{\mathbb{R}^3} f^n\phi(v)\,dv - \Delta t \sum_{j=1}^{i-1} \widetilde{a}_{ij} \int_{\mathbb{R}^3} v\cdot\nabla_x F^{(j)}\phi(v)\,dv \tag{102}$$

$$\int_{\mathbb{R}^3} f^{n+1}\phi(v)\,dv = \int_{\mathbb{R}^3} f^n\phi(v)\,dv - \Delta t \sum_{i=1}^{\nu} \widetilde{\omega}_i \int_{\mathbb{R}^3} v\cdot\nabla_x F^{(i)}\phi(v)\,dv. \tag{103}$$

Thus if (101) holds true the higher order moments of the $F^{(i)}$ can be computed as function of mass, momentum and temperature of $F^{(i)}$ and we get the explicit Runge-Kutta scheme applied to the corresponding system of compressible Euler equations. We can state:

Theorem 4 *If* $\det A \neq 0$, *in the limit* $\varepsilon \to 0$, *the IMEX scheme (98)-(99) applied to system (97) becomes the explicit RK scheme characterized by* $(\widetilde{A}, \widetilde{w}, \widetilde{c})$ *applied to the limit Euler system (23).*

Note however that the above results do not guarantee that

$$\lim_{\varepsilon\to 0} f^{n+1} = M[f^{n+1}].$$

The latter property is achieved, for example, if we assume that

$$\widetilde{a}_{\nu i} = \widetilde{w}_i, \quad a_{\nu i} = w_i, \quad i = 1, \ldots, \nu \tag{104}$$

with $a_{\nu\nu} \neq 0$. In fact, as a consequence we have $f^{n+1} = F^{(\nu)}$, and $\lim_{\varepsilon \to 0} F^{(\nu)} = M[F^{(\nu)}]$. As for the corresponding DIRK methods such schemes are referred to as *stiffly accurate*.

Let us remark that the IMEX scheme (98)-(99) can be solved explicitly despite the nonlinearity of $M[f]$. In fact since $M[F^{(i)}]$ depends only on mass, momentum and temperature of the solution, thus on the low order moments of $F^{(i)}$, it can be evaluated directly from the explicit scheme (102)-(103). This property has been used, for example, in Pieraccini and Puppo (2007); Filbet and Jin (2010) to implement efficiently IMEX methods.

Several AP IMEX schemes up to third order can be found in Pareschi and Russo (2005, 2000a). In Tables 9 and 10 we report two examples of second order methods, PR(2,2,2) method, satisfying $\widetilde{w} = w$ and $\det(A) \neq 0$, and ARS(2,2,2), satisfying $\widetilde{c} = c$ and the stiffly accurate requirement. A modification of such schemes in order to achieve uniformly accuracy in ε has been studied in Boscarino and Russo (2009).

4.3.2 Application to the Boltzmann equation

In the sequel we will describe how to extend the IMEX Runge-Kutta methods to the full Boltzmann equation (1). We follow the methodology introduced in Filbet and Jin (2010); Dimarco and Pareschi (2011).

Our scope is to avoid the implicit solution of the collision term of the Boltzmann equation. To this goal we can reformulate the collision operator adding a BGK (or another linear kinetic model which can be easily inverted) as a penalization, exactly as shown in Section 3.2.1. Let us mention that a similar approach has been previously used in other contexts (see, for example, (Smereka, 2003)) with the goal to adopt a time step which is much larger than $O(\varepsilon)$ in regions close to the fluid dynamic limit.

The Boltzmann equation can be rewritten as

$$\partial_t f + v \cdot \nabla_x f = \frac{\mu}{\varepsilon} g(f) + \frac{\mu}{\varepsilon}(M[f] - f), \tag{105}$$

with

$$\mu g(f) = P(f,f) - \mu M[f], \quad P(f,f) = Q(f,f) + \mu f.$$

Clearly now $M[f]$ is non constant in time during the relaxation process. One can now use a numerical scheme in which only the BGK term $M[f] - f$ is treated implicitly, while the term $g(f)$ describing the deviations from a BGK behavior and the convection term $\nabla_x f$ are treated explicitly.

The IMEX scheme now take the form

$$F^{(i)} = f^n + \Delta t \sum_{j=1}^{i-1} \widetilde{a}_{ij} \left(\frac{\mu}{\varepsilon} g(F^{(j)}) - v \cdot \nabla_x F^{(j)} \right) + \Delta t \sum_{j=1}^{\nu} a_{ij} \frac{\mu}{\varepsilon} (M[F^{(j)}] - F^{(j)}) \tag{106}$$

$$f^{n+1} = f^n + \Delta t \sum_{i=1}^{\nu} \widetilde{\omega}_i \left(\frac{\mu}{\varepsilon} g(F^{(i)}) - v \cdot \nabla_x F^{(i)} \right) + \Delta t \sum_{j=1}^{\nu} \omega_i \frac{\mu}{\varepsilon} (M[F^{(i)}] - F^{(i)}). \tag{107}$$

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array} \qquad \begin{array}{c|cc} \gamma & \gamma & 0 \\ 1-\gamma & 1-2\gamma & \gamma \\ \hline & 1/2 & 1/2 \end{array} \qquad \gamma = 1 - \frac{1}{\sqrt{2}}$$

Table 9: Tableau for the explicit (left) implicit (right) PR(2,2,2) scheme

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ 1 & \delta & 1-\delta & 0 \\ \hline & \delta & 1-\delta & 0 \end{array}, \quad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & 0 & \gamma & 0 \\ 1 & 0 & 1-\gamma & \gamma \\ \hline & 0 & 1-\gamma & \gamma \end{array}, \quad \gamma = 1 - \frac{\sqrt{2}}{2}, \delta = 1 - \frac{1}{2\gamma}$$

Table 10: Tableau for the explicit (left) implicit (right) ARS(2,2,2) scheme

If we integrate the above scheme against $\phi(v) = 1, v, v^2$ we obtain again system (102)-(103). It is a simple exercise to verify that Lemma 4 and Theorem 4 apply also to this reformulated problem. Thus we obtain AP method for the Boltzmann equation that can be evaluated explicitly. Here, however, the additional difficulty is given by the fact that we are integrating explicitly the stiff term $\mu g(f)/\varepsilon$ which may lead to unstable schemes if not properly treated. The stiffly accurate conditions (104) are essential in such a situation in order to obtain unconditionally stable AP schemes.

Finally let us emphasize that the usual stability requirements may not suffice to guarantee nonnegativity of the solution f^{n+1} starting from a nonnegative initial data f^n. We refer to Dimarco and Pareschi (2011) for further details.

Acknowledgements

The authors are grateful to G. Dimarco, F. Filbet, C. Mouhot and P. Santagati who contributed in a fundamental way to this research survey.

References

Armbruster, D., Degond, P., and Ringhofer, C. (2007). Kinetic and fluid models for supply chains supporting policy attributes. *Bulletin of the Institute of Mathematics*, 2:433–460.

Bellomo, N. and Bellouquid, A. (2004). From a class of kinetic models to the macroscopic equations for multicellular systems in biology. *Discrete Contin. Dyn. Syst. Ser. B*, 4:59–80.

Bennoune, M., Lemou, M., and Mieussens, L. (2008). Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier-Stokes asymptotics. *J. Comp. Phys.*, 227:3781–3803.

Bhatnagar, P., Gross, E., and Krook, M. (1954). A model for collision processes in gases i. small amplitute processes in charged and neutral one component systems. *Phys. Rev.*, 94:511–525.

Bird, G. (1994). *Molecular gas dynamics and direct simulation of gas flows.* Clarendon Press, Oxford.

Bobylev, A. (1975). Exact solutions of the Boltzmann equation (russian). *Dokl. Akad. Nauk. S.S.S.R.*, 225:1296–1299.

Bobylev, A. (1988). The theory of the nonlinear spatially uniform Boltzmann equation for maxwell molecules. *Math. Phys. Reviews*, 7:111–233.

Bobylev, A., Carrillo, J., and Gamba, I. (2000). On some properties of kinetic and hydrodynamics equations for inelastic interactions. *J. Statist. Phys.*, 98:743–773.

Bobylev, A., Palczewski, A., and Schneider, J. (1995). On approximation of the Boltzmann equation by discrete velocity models. *C. R. Acad. Sci. Parais Sér. I. Math.*, 320:639–644.

Bobylev, A. and Rjasanow, S. (1997). Difference scheme for the Boltzmann equation based on the fast fourier transform. *European J. Mech. B Fluids*, 16:293–306.

Bobylev, A. and Rjasanow, S. (1999). Fast deterministic method of solving the Boltzmann equation for hard spheres. *Eur. J. Mech. B Fluids*, 18:869–887.

Bobylev, A. and Rjasanow, S. (2000). Numerical solution of the Boltzmann equation using a fully conservative difference scheme based on the fast Fourier transform. *Transport Theory Statist. Phys.*, 29(3-5):289–310.

Boscarino, S. and Russo, G. (2009). On a class of uniformily accurate IMEX Runge-Kutta schemes and applications to hyperbolic systems with relaxation. *SIAM J. Sci. Comp*, 31:1926–1945.

Botchorishvili, R., Perthame, P., and Vasseur, A. (2003). Equilibrium schemes for scalar conservation laws with stiff sources. *Math. Comp.*, 72(241):131–157 (electronic).

Bouchut, F. and Perthame, B. (1993). A BGK model for small prandtl numbers in the navier-stokes approximation. *J. Stat. Phys.*, 71:191–207.

Bourgat, F., LeTallec, P., Perthame, B., and Qiu, Y. (1992). *Coupling Boltzmann and Euler equations without overlapping.* AMS, Providence, RI.

Bourgat, J., LeTallec, P., and Tidriri, M. (1996). Coupling Boltzmann and navier-stokes equations by friction. *J. Comput. Phys.*, 127:227–245.

Buet, C. (1996). A discrete velocity scheme for the Boltzmann operator of rarefied gas dynamics. *Trans. Theo. Stat. Phys.*, 25:33–60.

Buet, C., Cordier, S., Degond, P., and Lemou, M. (1997). Fast algorithms for numerical, conservative, and entropy approximations of the Fokker-Planck equation. *J. Comp. Phys.*, 133:310–322.

Caflisch, R. (1980). The fluid dynamical limit of the nonlinear Boltzmann equation. *Commun. Pure Appl. Math.*, 33:651–666.

Caflisch, R., Jin, S., and Russo, G. (1997). Uniformly accurate schemes for hyperbolic systems with relaxation. *SIAM J. Numer. Anal.*, 34:246–281.

Cai, Z. and Li, R. (2010). An h-adaptive mesh method for Boltzmann-BGK/hydrodynamics coupling. *J. Comput. Phys.*, 229(5):1661–1680.

Canuto, C., Hussaini, M., Quarteroni, A., and Zang, T. (1988). *Spectral methods in fluid dynamics.* Springer Series in Computational Physics, Springer-Verlag, New York.

Carleman, T. (1932). Sur la théorie de l'équation intégrodifférentielle de Boltzmann. *Acta Math.*, 60.

Carrillo, J., Gamba, I., Majorana, A., and Shu, C.-W. (2006). 2D semiconductor device simulations by WENO-Boltzmann schemes: efficiency, boundary conditions and comparison to Monte Carlo methods. *J. Comput. Phys.*, 214(1):55–80.

Cercignani, C. (1988). *The Boltzmann equation and its applications.* Springer Verlag, New York.

Cercignani, C., Illner, R., and Pulvirenti, M. (1994). *The mathematical theory of dilute gases*, volume 106. Applied Mathematical Sciences.

Cercignani, C. and Lampis, M. (1971). Kinetic models for gas-surface interactions. *Transport Th. Stat. Phys.*, 1:101–114.

Chalub, F., Markowich, P., Perthame, B., and Schmeiser, C. (2004). Kinetic models for chemotaxis and their drift-diffusion limits. *Monatsh. Math.*, 142:123–141.

Chorin, A. (1972). Numerical solution of Boltzmann's equation. *Comm. Pure Appl. Math.*, pages 171–186.

Cockburn, B., Johnson, C., Shu, C.-W., and Tadmor, E. (1998). *Advanced numerical approximation of nonlinear hyperbolic equations*, volume 1697 of *Lecture Notes in Mathematics.* Springer-Verlag, Berlin. Papers from the C.I.M.E. Summer School held in Cetraro, June 23–28, 1997, Edited by Alfio Quarteroni, Fondazione C.I.M.E.. [C.I.M.E. Foundation].

Cooley, J. and Tukey, J. (1965). An algorithm for the machine calculation of complex Fourier series. *Math. Comput.*, 19:297–301.

Cordier, S., Pareschi, L., and Toscani, G. (2005). On a kinetic model for a simple market economy. *J. Stat. Phys.*, 120:253–277.

Coron, F. and Perthame, B. (1991). Numerical passage from kinetic to fluid equations. *SIAM J. Numer. Anal.*, 28:26–42.

Degond, P., Dimarco, G., and Mieussens, L. (2007). A moving interface method for dynamic kinetic-fluid coupling. *J. Comput. Phys.*, 227(2):1176–1208.

Degond, P., Jin, S., and Mieussens, L. (2005). A smooth transition between kinetic and hydrodynamic equations. *J. Comp. Phys.*, 209:665–694.

Degond, P., Pareschi, L., and Russo, G. (2004). *Modeling and Computational Methods for Kinetic Equations.* Series: Modeling and Simulation in Science, Engineering and Technology. Birkhauser.

Desvillettes, L., Ferriere, R., and Prevost, C. (2004). Infinite dimensional reaction-diffusion for population dynamics. *(preprint).*

Desvillettes, L. and Mischler, S. (1996). About the splitting algorithm for boltzmann and BGK equations. *Math. Mod. & Meth. in App. Sci.*, 6:1079–1101.

Dia, B. and Schatzman, M. (1996)). Commutateurs de certains semi-groupes holomorphes et applications aux directions alternées. *M2AN Math. Model. Num. Anal.*, 30:343–383.

Dimarco, G. and Pareschi, L. (2010a). Exponential runge-kutta methods for stiff kinetic equations. *(Preprint).*

Dimarco, G. and Pareschi, L. (2010b). Fluid solver independent hybrid methods for multiscale kinetic equations. *SIAM J. Sci. Comput.*, 32(2):603–634.

Dimarco, G. and Pareschi, L. (2011). Implicit-explicit Runge-Kutta schemes for nonlinear kinetic equations. *preprint.*

Escobedo, M., Laurençot, P., Mischler, S., and Perthame, B. (2003a). Gelation and mass conservation in coagulation-fragmentation models. *J. Differential Equations*, 195(1):143–174.

Escobedo, M., Mischler, S., and Valle, M. D. (2003b). Homogeneous Boltzmann equation for quantum and relativistic particles. *Electron. J. Diff. Eqns.*, Monograph 04:85 pages.

Filbet, F., Hu, J., and Jin, S. (2011). A numerical scheme for the quantum Boltzmann equation efficient in the fluid regime. *M2AN Math. Model. Numer. Anal.*, to appear.

Filbet, F. and Jin, S. (2010). A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources. *J. Comput. Phys.*, 229:7625–7648.

Filbet, F. and Mouhot, C. (2011). Analysis of spectral methods for the homogeneous Boltzmann equation. *Trans. Amer. Math. Soc.*, 363:1947–1980.

Filbet, F., Mouhot, C., and Pareschi, L. (2006). Solving the Boltzmann equation in $o(n \log n)$. *SIAM J. Sci. Comput.*, 28:1029–1053.

Filbet, F. and Pareschi, L. (2003). A numerical method for the accurate solution of the Fokker-Planck-Landau equation in the non homogeneous case. *J. Comput. Phys.*, 186:457–480.

Filbet, F., Pareschi, L., and Toscani, G. (2005). Accurate numerical methods for the collisional motion of (heated) granular flows. *J. Comput. Phys.*, 202:216–235.

Filbet, F. and Russo, G. (2003). High order numerical methods for the space non-homogeneous Boltzmann equation. *J. Comput. Phys.*, 186:457–480.

Filbet, F. and Russo, G. (2006). A rescaling velocity method for kinetic equations: the homogeneous case. In *Modelling and numerics of kinetic dissipative systems*, pages 191–202. Nova Sci. Publ., Hauppauge, NY.

Filbet, F., Sonnendrücker, E., and Bertrand, P. (2001). Conservative numerical schemes for the vlasov equation. *J. Comput. Phys.*, 172:166–187.

Gabetta, E., Pareschi, L., and Toscani, G. (1997). Relaxation schemes for nonlinear kinetic equations. *SIAM J. Numer. Anal.*, 34:2168–2194.

Gamba, I. and Tharkabhushanam, S. (2009). Spectral-Lagrangian methods for collisional models of non-equilibrium statistical states. *J. Comput. Phys.*, 228(6):2012–2036.

Gamba, I. and Tharkabhushanam, S. (2010). Shock and boundary structure formation by spectral-Lagrangian methods for the inhomogeneous Boltzmann transport equation. *J. Comput. Math.*, 28(4):430–460.

Golse, F. and Saint-Raymond, L. (2004). The navier-stokes limit of the Boltzmann equation for bounded collision kernels. *Invent. Math.*, 155:81–161.

Greenberg, J. and Leroux, A. (1996). A well-balanced scheme for the numerical processing of source terms in hyperbolic equations. *SIAM J. Numer. Anal.*, 33(1):1–16.

Ha, S.-. and Tadmor, E. (2008). From particle to kinetic and hydrodynamic descriptions of flocking. *Kinet. Relat. Models*, 1(3):415–435.

Hairer, E., Lubich, C., and Wanner, G. (2002). *Geometric Numerical Integration. Structure- Preserving Algorithms for Ordinary Differential Equations*. Springer, Berlin.

Hairer, E. and Wanner, G. (1996). *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, volume 14 of *Springer Series in Computational Mathematics*. Springer-Verlag, second revised edition.

Hardy, G. and Wright, E. (1979). *An introduction to the theory of numbers*. The Clarendon Press Oxford University Press, New York, fifth edition.

Heintz, A., Kowalczyk, P., and Grzhibovskis, R. (2008). Fast numerical method for the Boltzmann equation on non-uniform grids. *J. Comput. Phys.*, 227(13):6681–6695.

Higueras, I. (2005). Representations of runge-kutta methods and strong stability preserving methods. *SIAM J. Numer. Anal.*, 43:924–948.

Holway, L. (1966). New statistical models for kinetic theory: methods of construction. *Phys. Fluid.*, 9:1658–1673.

Ibragimov, I. and Rjasanow, S. (2002). Numerical solution of the Boltzmann equation on the uniform grid. *Computing*, 69:163–186.

Jin, S. (1995). Runge-kutta methods for hyperbolic conservation laws with stiff relaxation terms. *J. Comp. Phys.*, 122:51–67.

Jin, S., Pareschi, L., and Toscani, G. (2000). Uniformly accurate diffusive relaxation schemes for multiscale transport equations. *SIAM J. Numer. Anal.*, 38:913–936.

Kac, M. (1957). *Probability and related topics in the physical sciences.* Interscience Publishers.

Klar, A. and Wegener, R. (1997). Enskog-like kinetic models for vehicular traffic. *J. Stat. Phys.*, 87:91–114.

Landau, L. (1981). *The transport equation in the case of the Coulomb interaction*, pages 163–170. D.ter Haar Ed. Collected papers of L.D. Landau. Pergamon press.

Lemou, M. (1998). Multipole expansions for the Fokker-Planck-Landau operator. *Num. Math.*, 78:597–618.

Lemou, M. and Mieussens, L. (2008). A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit. *SIAM J. Sci. Comput.*, 31(1):334–368.

LeVeque, R. (1992). *Numerical methods for conservation laws.* Birkhauser Verlag.

Levermore, C. (1996). Moment closure hierarchies for kinetic theories. *J. Stat. Phys.*, 83:1021–1065.

Markowich, P. and Pareschi, L. (2005). Fast conservative and entropic numerical methods for the boson Boltzmann equation. *Num. Math.*, 99:509–532.

Markowich, P., Ringhofer, C., and Schmeiser, C. (1989). *Semiconductor equations.* Springer-Verlag.

Martin, Y.-L., Rogier, F., and Schneider, J. (1992). Une méthode déterministe pour la résolution de l'équation de Boltzmann inhomogène. *C. R. Acad. Sci. Paris Sér. I Math.*, 314:483–487.

Maset, S. and Zennaro, M. (2009). Unconditional stability of explicit exponential runge-kutta methods for semi-linear ordinary differential equations. *Math. Comp.*, 78:957–Ű967.

McLachlan, R. (1995). On the numerical integration of ordinary differential equations by symmetric composition methods. *Siam J. Sci. Comp.*, 16:151–168.

Mieussens, L. (2000). Discrete velocity model and implicit scheme for the BGK equation of rarefied gas dynamics. *Math. Models and Meth. Appl. Sci.*, 8:1121–1149.

Mouhot, C. and Pareschi, L. (2004). Fast methods for the Boltzmann collision integral. *C. R. Math. Acad. Sci. Paris*, 339(1):71–76.

Mouhot, C. and Pareschi, L. (2006). Fast algorithms for computing the Boltzmann collision operator. *Math. Comp.*, 75(256):1833–1852 (electronic).

Mouhot, C. and Pareschi, L. (2011). An $o(n \log n)$ algorithm for computing discrete velocity models. *(preprint).*

Naldi, G., Pareschi, L., and Toscani, G. (2003). Spectral methods for one-dimensional kinetic models of granular flows and numerical quasi elastic limit. *ESAIM RAIRO Math. Model. Numer. Anal.*, 37:73–90.

Naldi, G., Pareschi, L., and Toscani, G. (2010). *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences*. Birkhauser, Boston.

Nanbu, K. (1980). Direct simulation scheme derived from the Boltzmann equation i. monocomponent gases. *J. Phys. Soc. Japan*, 49:2042–2049.

Nishida, T. (1978). Fluid dynamical limit of the nonlinear Boltzmann equation at the level of the compressible euler equations. *Commun. Math. Phys.*, 61:119–148.

Ohwada, T. (1993). Structure of normal shock waves: Direct numerical analysis of the Boltzmann equation for hard sphere molecules. *Phys. Fluids A*, 5:217–234.

Palczewski, A. and Schneider, J. (1998). Existence, stability, and convergence of solutions of discrete velocity models to the Boltzmann equation. *J. Statist. Phys.*, 91:307–326.

Palczewski, A., Schneider, J., and Bobylev, A. (1997). A consistency result for a discrete-velocity model of the Boltzmann equation. *SIAM J. Numer. Anal.*, 34:1865–1883.

Panferov, V. and Heintz, A. (2002). A new consistent discrete-velocity model for the Boltzmann equation. *Math. Methods Appl. Sci.*, 25:571–593.

Pareschi, L. and Perthame, B. (1996). A spectral method for the homogeneous Boltzmann equation. *Trans. Theo. Stat. Phys.*, 25:369–383.

Pareschi, L. and Russo, G. (2000a). Implicit-explicit Runge-Kutta schemes for stiff system of differential equations. *Recent Trend in Numerical Analysis*, 3:269–289.

Pareschi, L. and Russo, G. (2000b). Numerical solution of the Boltzmann equation i. spectrally accurate approximation of the collision operator. *SIAM J. Numer. Anal.*, 37:1217–1245.

Pareschi, L. and Russo, G. (2000c). On the stability of spectral methods for the homogeneous Boltzmann equation. *Trans. Theo. Stat. Phys.*, 29:431–447.

Pareschi, L. and Russo, G. (2005). Implicit-explicit runge-kutta methods and applications to hyperbolic systems with relaxation. *J. Sci. Comp.*, 25:129–155.

Pareschi, L., Russo, G., and G.Toscani (2000). Fast spectral methods for the Fokker-Planck-Landau collision operator. *J. Comput. Phys.*, 165:216–236.

Pareschi, L., Toscani, G., and Villani, C. (2003). Spectral methods for the non cut-off Boltzmann equation and numerical grazing collision limit. *Numer. Math.*, 93(3):527–548.

Pieraccini, S. and Puppo, G. (2007). Implicit-Explicit schemes for BGK kinetic equations. *Journal of Scientific Computing*, 32:1–28.

Platkowski, T. and Illner, R. (1988). Discrete velocity models of the Boltzmann equation: A survey on the mathematical aspects of the theory. *SIAM Review*, 30(2):213–255.

Roe, P. and Sidilkover, D. (1992). Optimum positive linear schemes for advection in two and three dimensions. *SIAM J. Numer. Anal.*, 29:1542–1568.

Rogier, F. and Schneider, J. (1994). A direct method for solving the Boltzmann equation. *Trans. Theo. Stat. Phys.*, 23:313–338.

Russo, G. and Santagati, P. (2011). A new class of conservative large time step methods for the BGK models of the Boltzmann equation. *preprint*, arXiv:1103.5247.

Santagati, P. (2007). *High order semi-Lagrangian schemes for the BGK model of the Boltzmann equation*. Department of Mathematics and Computer Science, University of Catania. PhD. thesis.

Schwartzentruber, T., Scalabrin, L., and Boyd, I. (2007). A modular particle-continuum numerical method for hypersonic non-equilibrium gas ows. *Journal of Computational Physics*, 225:1159–1174.

Smereka, P. (2003). Semi-implicit level set methods for curvature and surface diffusion motion. *J. Sci. Comput.*, 19:439–456.

Sod, G. (1977). A numerical solution of Boltzmann's equation. *Comm. Pure Appl. Math.*, 30:391–419.

Sone, Y., Aoki, K., Takata, S., Sugimoto, H., and Bobylev, A. (1996). Inappropriateness of the heat-conduction equation for description of a temperature field of a stationary gas in the continuum limit: examination by asymptotic analysis and numerical computation of the Boltzmann equation. *Phys. Fluids*, 8:628–638.

Sone, Y., Ohwada, T., and Aoki, K. (1989). Temperature jump and knudsen layer in a rarefied gas over a plane wall: Numerical analysis of the linearized Boltzmann equation for hard-sphere molecules. *Phys. Fluids A*, 1:363–370.

Strang, G. (1968). On the construction and the comparison of difference schemes. *SIAM J. Numer. Anal.*, 5:506–517.

Tiwari, S. (1998). Coupling of the Boltzmann and euler equations with automatic domain decomposition. *J. Comput. Phys.*, 144:710–726.

Tiwari, S. and Klar, A. (1998). An adaptive domain decomposition procedure for Boltzmann and euler equations. *J. Comp. Appl. Math.*, 90:223–237.

Villani, C. (2002). *A survey of mathematical topics in kinetic theory*. Handbook of fluid mechanics, S. Friedlander and D. Serre, Eds. Elsevier Publ.

Whitham, G. (1974). *Linear and nonlinear waves*. J.Wiley.